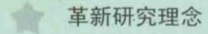


格致方法·定量研究系列 吴晓刚 主编

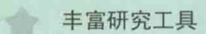
logit与probit：次序模型和多类别模型

[英] 瓦尼·布鲁雅（Vani Borooah）著

张卓妮 译



革新研究理念



丰富研究工具

最权威、最前沿的定量研究方法指南

14

格致出版社



格致方法·定量研究系列

1. 社会统计的数学基础
2. 理解回归假设
3. 虚拟变量回归
4. 多元回归中的交互作用
5. 回归诊断简介
6. 现代稳健回归方法
7. 固定效应回归模型
8. 用面板数据做因果分析
9. 多层次模型
10. 分位数回归模型
11. 空间回归模型
12. 删截、选择性样本及截断数据的回归模型
13. 应用logistic回归分析（第二版）
14. logit与probit：次序模型和多类别模型
15. 定序因变量的logistic回归模型
16. 对数线性模型
17. 流动表分析
18. 关联模型
19. 中介作用分析
20. 因子分析：统计方法与应用问题
21. 非递归因果模型
22. 评估不平等
23. 分析复杂调查数据（第二版）
24. 分析重复调查数据
25. 世代分析
26. 纵贯研究
27. 多元时间序列模型
28. 潜变量增长曲线模型
29. 缺失数据
30. 社会网络分析
31. 广义线性模型导论
32. 基于行动者的模型
33. 基于布尔代数的比较法导论
34. 微分方程：一种建模方法
35. 模糊集合理论在社会科学中的应用
36. 图形代数
37. 项目功能差异

上架建议：社会研究方法
ISBN 978-7-5432-2128-4



定价：15.00元
易文网：www.ewen.c
格致网：www.hibooks

格致方法·定量研究系列 吴晓刚 主编

logit与probit：次序模型和多类别模型

[英] 瓦尼·布鲁雅(Vani Borooah) 著
张卓妮 译

SAGE Publications ,Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

logit与probit:次序模型和多类别模型/(英)布鲁雅(Borooah, V.)著;张卓妮译. —上海:格致出版社:上海人民出版社,2012
(格致方法·定量研究系列)
ISBN 978-7-5432-2128-4

Ⅰ. ①l… Ⅱ. ①布… ②张… Ⅲ. ①社会科学-定量方法-研究 Ⅳ. ①C34

中国版本图书馆CIP数据核字(2012)第130319号

责任编辑 高 璇

格致方法·定量研究系列
logit与probit:次序模型和多类别模型
[英]瓦尼·布鲁雅 著
张卓妮 译

出 版 世纪出版集团 www.ewen.cc 格致出版社 www.hibooks.cn 上海人民出版社
(200001 上海福建中路193号24层)


编辑部热线 021-63914988
市场部热线 021-63914081

发 行 世纪出版集团发行中心
印 刷 浙江临安曙光印务有限公司
开 本 920×1168毫米 1/32
印 张 4.5
字 数 85,000
版 次 2012年7月第1版
印 次 2012年7月第1次印刷
ISBN 978-7-5432-2128-4/C·78
定 价 15.00元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书，精选了世界著名的 SAGE 出版社定量社会科学研究丛书中的 35 种，翻译成中文，集结成八册，于 2011 年出版。这八册书分别是：《线性回归分析基础》、《高级回归分析》、《广义线性模型》、《纵贯数据分析》、《因果关系模型》、《社会科学中的数理基础及应用》、《数据分析方法五种》和《列表数据分析》。这套丛书自出版以来，受到广大读者特别是年轻一代社会科学工作者的欢迎，他们针对丛书的内容和翻译都提出了很多中肯的建议。我们对此表示衷心的感谢。

基于读者的热烈反馈，同时也为了向广大读者提供更多的方便和选择，我们将该丛书以单行本的形式再次出版发行。在此过程中，主编和译者对已出版的书做了必要的修订和校正，还新增加了两个品种。此外，曾东林、许多多、范新光、李忠路协助主编参加了校订。今后我们将继续与 SAGE 出版社合作，陆续推出新的品种。我们希望本丛书单行本的出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

往事如烟，光阴如梭。转眼间，出国已然十年有余。1996年赴美留学，最初选择的主攻方向是比较历史社会学，研究的兴趣是中国的制度变迁问题。以我以前在国内所受的学术训练，基本是看不上定量研究的。一方面，我们倾向于研究大问题，不喜欢纠缠于细枝末节。国内一位老师的话给我的印象很深，大致是说：如果你看到一堵墙就要倒了，还用得着纠缠于那堵墙的倾斜角度究竟是几度吗？所以，很多研究都是大而化之，只要说得通即可。另一方面，国内（十年前）的统计教学，总的来说与社会研究中的实际问题是相脱节的。结果是，很多原先对定量研究感兴趣的学生在学完统计之后，依旧无从下手，逐渐失去了对定量研究的兴趣。

我所就读的美国加州大学洛杉矶分校社会学系，在定量研究方面有着系统的博士训练课程。不论研究兴趣是定量还是定性的，所有的研究生第一年的头两个学期必须修两门中级统计课，最后一个学期的系列课程则是简单介绍线性回归以外的其他统计方法，是选修课。希望进一步学习定量研

究方法的可以在第二年修读另外一个三学期的系列课程，其中头两门课叫“调查数据分析”，第三门叫“研究设计”。除此以外，还有如“定类数据分析”、“人口学方法与技术”、“事件史分析”、“多层线性模型”等专门课程供学生选修。该学校的统计系、心理系、教育系、经济系也有一批蜚声国际的学者，提供不同的、更加专业化的课程供学生选修。2001年完成博士学业之后，我又受安德鲁·梅隆基金会资助，在世界定量社会科学研究的重镇密歇根大学从事两年的博士后研究，其间旁听谢宇教授为博士生讲授的统计课程，并参与该校社会研究院(Institute for Social Research)定量社会研究方法项目的一些讨论会，受益良多。

2003年，我赴港工作，在香港科技大学社会科学部，教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来，非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课，而第二门课“社会科学中的定量分析”为博士生的必修课(事实上，大部分硕士生在修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候，根据社会科学研究生的数理基础比较薄弱的特点，尽量避免复杂的数学公式推导，而用具体的例子，结合语言和图形，帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上，即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”，学完这些课程后，我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章；另一方面，也能在自己的研究中运用这些成熟的

方法论技术。

上述两门课的内容，尽管在线性回归模型的内容上有少量重复，但各有侧重。“社会科学里的统计学”(Statistics for Social Science)从介绍最基本的社会研究方法论和统计学原理开始，到多元线性回归模型结束，内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型，以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法，将重点放在因变量为定类数据的分析模型上，包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型，包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生，一直鼓励和支持我将两门课的讲稿结集出版，并帮助我将原来的英文课程讲稿译成了中文。但是，由于种种原因，这两本书拖了四年多还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩，每本书都写得通俗易懂。中山大学马骏教授向格致出版社何元龙社长推荐了这套书，当格致出版社向我提出从这套丛书中精选一批翻译，以飨中文读者时，我非常支持这个想法，因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种

语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及港台地区的二十几位研究生参与了这项工程,他们目前大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是:

香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾"中央"研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦。

关于每一位译者的学术背景,书中相关部分都有简单的介绍。尽管每本书因本身内容和译者的行文风格有所差异,校对也未免挂一漏万,术语的标准译法方面还有很大的改进空间,但所有的参与者都做了最大的努力,在繁忙的学习和研究之余,在不到一年的时间内,完成了三十五本书、超过百万字的翻译任务。李骏、叶华、张卓妮、贺光烨、宋曦、於嘉、郑冰岛和林宗弘除了承担自己的翻译任务之外,还在初稿校对方面付出了大量的劳动。香港科技大学霍英东南沙研究院的工作人员曾东林,协助我通读了全稿,在此

我也致以诚挚的谢意。有些作者，如香港科技大学黄善国教授、美国约翰·霍普金斯大学郝令昕教授，也参与了审校工作。

我们希望本丛书的出版，能为建设国内社会科学定量研究的扎实学风作出一点贡献。

吴晓刚

于香港九龙清水湾

序

要使普通最小二乘法(OLS)产生最优线性无偏估计(BLUE),必须符合经典回归假设。这些假设中有些假设比其他假设更容易实现。此外,违反这些假设的实际后果因假设的不同而不同。其中一个假设难以实现,而且会对OLS的解释造成严重后果,那就是假设因变量是连续的。相反,如果因变量是离散的,即由两个或更多的结果类别构成,那么OLS就会产生严重的推论问题。在这种情况下,最大似然(maximum likelihood)技术如logit或probit通常更有效。

本书比较独特,因为它完全致力于分析当因变量具多类别时的估计情况。在概论之后,作者关注了具离散和次序形式的因变量。比如,假设某位政治科学家有选举调查的数据,并希望解释政治兴趣这一因变量,其中受访者的得分:0 = 低,1 = 中等,2 = 高。这个变量是离散的,受访者处于这三种类别中的一种。此外,这个变量是从“低”到“高”排序的。在这种有序变量情况下,我们可以说某个得分为“高”的人比某个得分为“低”的人具有更多的政治兴趣,但我们不能确切地说多多少。所以,OLS回归看起来较不可取,而次序logit

或次序 probit 更可取,因为它们适合这种较低的测量水平。布鲁雅(Borooah)教授详尽地阐释了这两种方法,试图解释社会剥夺(用三个类别测量,"没有被剥夺"、"轻度被剥夺"、"严重被剥夺")在不同个体间的差异。一个经常出现的问题是 logit 是否比 probit 更优,或者反之。这两种方法根本上的理论差异涉及误差项的分布是逻辑分布还是正态分布。实际上,正如本书指出的,我们很难提供足够的理由说明为什么选择其中一种方法而非另外一种。

本书还把处理对象扩展到具两个以上结果的多类别或非次序因变量。比如,宗教的选择、住宅区的选择、购物中心的选择、工作的选择等。多类别 logit 的一个关键假设是无关选项独立性(IIA)。正如布鲁雅教授所论述的,这个假设既是此技术的优点又是其缺点。他还对比数比(odds-ratios)和风险比(risk-ratios)做了重要但往往被忽视的区分。在二分类 logit 中,这两种比率之间没有差异,但是,在多类别 logit 中,结果是以风险比的方式显示的。

本书结尾给出了非常有用的计算机程序详情,用于说明书中的表格结果是如何产生的。这种逐步对计算机程序进行注释的方式让读者明白如何运行数据分析。讲解中具体使用的软件是 STATA,但作者还指出了 SAS, SPSS 和 LIMDEP 中其他可用的程序。总体而言,这本书为估计和解释从更复杂的离散因变量模型中得到的结果提供了一个有用的指南。

——迈克尔·刘易斯-贝克

Logit and Probit: Ordered and Multinmial Models

上海市版权局著作权合同登记号:图字 09-2009-552

目 录

第1章

概　论

戏剧评论家肯尼思·泰南(Kenneth Tynan)曾出色地描述过他的职业:由“知道路线,但是不知道怎么开车”的人组成。社会科学研究者经常发现自己处于类似的情形之中。有些人能在研究蓝图中定位,却不知如何从一个地方到另一个地方。其他人可以在研究中驱车前进却无法超越固定的路线。只有少数幸运的人既能导航又能飞行。对这种现象的察觉界定了本书的大概意图:(1)探寻穿越次序和多类别logit模型“地形”的主要路径;(2)为在这些路径中驱车前进提供指引。但首先,在发动机器之前,我们先对其简单介绍一下。

很多时候,适合我们分析的变量仅仅只是对定性结果的一种编码。这些模型就是定性选择模型。举例来说,在评判某种政府行为时,一个人非常赞成(编码 = 1),赞成(编码 = 2),反对(编码 = 3),或者强烈反对(编码 = 4);或者一个人投自由党的票(编码 = 1),投保守党的票(编码 = 2),或者工党的票(编 码 = 3)。在这种情形之下,因变量取的是一组离散的、互斥的、完全穷尽的值。这与其他至少在理论上假设因变量的取值是连续数值的情况不同。虽然传统的回归方法不适合离散因变量的统计分析,但是我们可以用回归分析

的原理，建立把观察到的结果与特定“决定性的”或“解释性的”变量连接起来的模型。因变量取值大于两个的定性选择模型为多元结果模型。我们还可以进一步把多元结果划分为：次序结果（如前述对政府行为的赞成程度的例子）和非次序结果（如前述投票的例子）。

定性选择模型在应用计量经济学分析中已经很普遍。社会科学家也一直对在互斥的选项中进行选择的问题感兴趣。越来越多可利用的调查数据（横截面或面板格式）意味着越来越多的学者可以把他们的知识思索转变成实实在在的成果。学者们带着一系列具体的问题从这些数据中寻找答案。然后，当分析完这些数据之后，结果往往隐藏在关于什么才是“正确”答案的不确定性之中。为了改善这种状况，必须对教科书技术进行巧妙的处理，以使推导出的结果明确地指向正确的方向。本章在社会经济不平等相关问题的背景之下，讨论了对多元结果模型——包括次序和非次序结果——的估计、模拟和解释。

次序和非次序模型对各自的分析有不同的技术要求。次序模型可以由 logit 方法即次序 logit 模型，或者 probit 方法即次序 probit 模型进行估计。非次序结果模型由 logit 方法估计最方便，虽然原则上用 probit 的方法也可以对这种模型进行估计，但是由于计算方面的原因，常常并不可行。因此，非次序多元结果模型被称为多类别 logit 模型。

多类别 logit 模型可以是有条件的，就是说在不同选项之间进行的选择不仅仅取决于做决定的个人所具有的特征，同时还取决于选项本身的属性。举个例子，个人选择惠顾哪个购物中心可能取决于购物中心本身的属性（商店的数量与种

类,中心的保养标准),这一属性并不随个人情况而变;同时也取决于收入和家庭大小,这些随个人而变。一个更复杂的因素是个人特征和选项属性之间还可能互相影响:对购物中心的选择可能决定于个人从居住地(居住地属于个人的特征)到某个特定的购物中心(购物中心的位置属于选项的属性)的距离。次序模型将在第 2 章中论述,多类别模型将在第 3 章中论述,这两章具有两个共同的目的:(1)介绍这些模型的基本方法论;(2)阐述在社会科学研究中如何运用这些模型。

满足第一个目的需要在过度单纯化与过分技术性之间寻找一个平衡点。作者已经尝试着找到这种平衡。然而,为了减轻负担,作者假设读者已经掌握了足够的知识。其中,这些知识包括:

(1) 了解线性概率模型的不足之处,或者为什么普通的回归分析并不适用于具离散因变量的模型。

(2) 熟悉因变量只有两个可能结果时的 logit 和 probit 方法。

关于以上这两点,这个系列里奥尔德里奇和纳尔逊(Aldrich & Nelson, 1984)、德马里斯(DeMaris, 1992)、廖(Liao, 1994)和梅纳德(Menard, 1995)等人的专著,以及更具概括性的计量经济学教材如格林(Greene)2000 年的著作,提供了非常好的综述。

满足第二个目的在作者看来更加困难。对一个模型的运用需要好几个认识层次。首先,我们必须非常清楚要提的问题。其次,如何回答它们也是一个问题。通常,一个特定的研究问题可以用不止一种方法来回答,所以知道这些不同

的方法怎么不同，为什么不同，以及眼前什么方法可能是最好的方法就很重要。最后，当我们已经决定了要提的问题以及怎么回答它们，就要处理另一个实践问题即如何通过取得并解释分析结果来执行我们的研究策略。

作者已经尝试着解决这些议题，即在次序 logit 和 probit，以及多类别 logit 模型的框架中采取一个三重策略。第一，作者尽可能把方法论和经验分析结合起来。几乎每一章的理论总结都可在随后的应用内容中找到对应的分析。

第二，作者试图用两个固定的应用项目来进行阐释。这样做是因为作者觉得用一个单一的经验（“真实世界”）线索贯穿全章比用相互脱节的例子分开阐释更有利于说明模型的应用。

第 2 章次序 logit 和 probit 以对社会剥夺状态的经验分析为基础。这个项目使用的数据来自北爱尔兰对将近 14000 个人进行的调查，该调查的主要内容是社会剥夺的根源以及天主教徒与新教徒之间被剥夺经历的不平等状况。第 3 章多类别 logit 模型基于对英国少数民族的职业获得的经验研究。这个项目使用的数据来自英国对将近 10 万全职男性雇员进行的普查，旨在研究黑加勒比人、印第安人和白人获得各种职业类别的不同机会。

第三，最后一章列举了用于产生经验结果的完整的电脑程序清单。据作者所知，至少有四种众所周知的具高度好评的软件，这些软件除了能解决其他方面的问题，还很适用于解决与本书的讨论内容类似的问题：SAS，SPSS v10.0（印第安纳大学统计与数学计算中心为在次序与非次序结果事件的分析中应用 SAS 和 SPSS 提供了很好的入门课程[1]），

LIMDEP(Greene，1995)以及 STATA(STATA，1999)。由于作者对 STATA 比较熟悉，因此第 4 章中的程序均用 STATA 编写。几乎每一行命令都附有一个注解，以说明这些命令应起的作用以及其如何与前面章节中的材料相关。

第2章

次序模型

第1节 | 简介

假设有 N 个人(标号为 $i=1, \cdots, N$),每个人都会发生一个"事件"。假设这个事件有 $M>2$ 种结果,标号为 $j=1, \cdots, M$,这些结果是相互排斥且完全穷尽的。令变量 Y_i 取的值代表第 i 个人的结果:$Y_i=1$ 表示这个人出现了第一种结果 ($j=1$),$Y_i=2$ 表示这个人出现了第二种结果 ($j=2$),如此这般,直至 $Y_i=M$ 表示出现最后一种结果 ($j=M$)。进一步假设这些结果在本质上就是次序的,意味着与变量 Y_i 更高的取值相关的结果比取值更低的结果排序更高。或者说,因变量 Y_i 与有序的结果相关:"更强的"结果与因变量更高的值联系在一起。但是,这些结果序号间的差异并不表示它们在程度上的差异;$Y_i=2$ 的结果并不是 $Y_i=1$ 的结果的两倍强度。因此,一个有序因变量的实际取值之间是不相关的,它只意味着更大的取值对应于更强的结果:我们可以令 $Y_i=5$ 表示发生了第一种结果,令 $Y_i=7$ 表示发生了第二种结果,以此类推。

一个人的健康状况是次序结果的例子。"差"、"良好"以及"非常好"的健康状况可以用一个取值分别为 1, 2 和 3 的变量表示。在这个例子中,$Y_i=3$(健康状况非常好)表示其结果比 $Y_i=2$(健康状况良好)更好,依次,后者又比 $Y_i=1$

(健康状况差)更好。其他次序结果的例子,如个人的保险范围(没有、半保、全保),或者处于工作年龄的个人之雇佣状况(不从事经济活动、失业、雇员)。而个人的宗教信仰则为非次序变量的例子,如 $Y_i = 1$ 表示新教徒,$Y_i = 2$ 表示犹太教徒,$Y_i = 3$ 表示穆斯林,$Y_i = 4$ 表示印第安教徒。虽然此例中的这些结果都不一样,但是它们却不可以排序,因此不能被视为有序结果。换一种方式说,这个表示宗教信仰的因变量是非次序的。

当结果明显具有次序的时候,我们必须注意:因变量既是离散的又是有序的。例如,若结果被编码为 1, 2, 3,线性回归就会把 3 与 2 之间的差异与 2 与 1 之间的差异同等对待,但其实这些数字其实只是一种排列方式而并没有什么重要意义。另一方面,用多类别 logit 的方法对一个有序因变量进行计量经济学关系的估计,则意味着这些数据的次序性所传达的信息被舍弃掉了。因此,用于估计因变量多于两个结果[2]且这些结果同时是离散的又是有序的模型,最常用且适当的方法就是次序 logit 和次序 probit 的方法[3]。

但是,以上所述受限于一个重要的附加说明。次序 logit 和 probit 的一个关键假设是平行斜率假设。这个假设的含义将在后面进行详细的讨论,但大体上,它的意思是如果有一个变量影响了一个人处于这些次序类别中的某一个结果的可能性(如饮食对健康状况的影响),就假定这个变量与结果间相关联的系数对所有的结果是一样的(特定的饮食对一个人处于非常好的健康状况的可能性的影响与对他/她处于差的健康状况的可能性的影响是一模一样的)。如果这个假设不成立,即,与某一特定变量相关联的斜率系数对不同的

结果是不一样的(特定的饮食对一个人处于非常好的健康状况的可能性的影响不同于对他/她处于差的健康状况的可能性的影响),那么次序 logit 和 probit 的方法就不再适当,这种模型就应该用多类别 logit 的方法。

我们并不是总能明确断定结果是有序的(还是无序的),这是另外一个需要谨慎使用次序估计方法的原因。例如,一个人居住于城市的哪个位置(北边,南边,东边或者西边)表面上看是一个非有序的变量。但是如果我们知道城市的某些位置相比其他位置具有更健康的居住条件,那么界定个人居住位置的变量就需要有序的内涵,比如,住在北边($Y_i = 4$)比住在南边($Y_i = 3$)要好。但是,当并不确定某个变量是次序的还是非次序的时候,一个明智的选择是把它当成非次序的来对待,并用多类别 logit 的方法对以它为因变量的模型进行估计。这个准则是切合实际的,因为如果把一个事实上是非次序的结果变量当成次序的来对待,就给这些其实无序的结果强加了一个排序并提出了平行斜率(参考上文,并于下文有详细讨论)的限制性假设,而这样很可能会使估计发生偏差。另一方面,如果把次序结果变量当成非次序的,未能对这些结果给予合理的排序,会损失估计的效率,但却不可能使估计发生偏差。在这两种可能出现的错误之中,效率的损失与有偏的估计相比并不那么严重。

这种不确定性的另一个例子,同时也是下一章多类别 logit 中的应用实例,是个人的职业。一个人从事的职业是非技术的、半技术的还是技术的;以及其工作是专业类别还是管理类别,都可以看做是一种个人选择,尽管影响这个选择的限制条件可能因人而异且尤其随种族和/或性别而改变。

在解释这种选择的时候，职业结果可以看做是“非次序的”，意即一种职业类别相比另一种职业类别并非固有地更理想或更不理想。对于这种解释，分析职业结果的适当的估计方法是多类别 logit。其实，这正是施米特和斯特劳斯（Schmid & Strauss，1975）对 1000 个人进行关于教育、经验、种族和性别的职业分析时所采用的方法。这种方法后来在格林（Greene，2000:859）讨论多类别 logit 时作为例子被应用。

从另一方面来说，假如一个大学教授被问及他/她更喜欢一个银行家还是一个看门人作为他/她的女婿，那个教授更可能会选择前者。在表达这种偏好时，这个教授暗中对职业进行了排序，把银行家排在了比看门人更高的层次（选择了更理想的女婿）。但是，重点是这种排序纯粹是主观的[4]（即银行家与看门人相比并不是固有地更理想），而且附带地，那个教授对于他/她的女婿应该健康而非不健康的偏好并不具备客观性。这个故事的寓意在于，我们最好把结果当成是非次序的来对待，除非有好的理由对这些结果强加一种排序。[5]

第 2 节 | 方法论

用一个具体的例子来说明次序 logit 和 probit 模型的方法论及其根本逻辑可能是最好的办法。假设有 N 个人(标号为 $i = 1, \cdots, N$),住在一个地区,且每个人的"社会剥夺程度"(degree of deprivation)可以用变量 D_i 的取值来表示,且 D_i 越高的取值代表越高的剥夺程度。对某个特定个人赋予的"剥夺指数"(deprivation index)假设值——此后称做他/她的"剥夺分数"——取决于这个人既有的多种因素。

这些因素举例来说可能包括失业、单亲家庭及居住于某一特定区域。假设剥夺指数 D_i,是 K 个因素(决定变量)的一个线性函数,这 K 个因素的取值对于个人 i 来说,为 X_{ik},$k = 1, \cdots, K$。这意味着剥夺指数可以表示为:

$$D_i = \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i = Z_i + \varepsilon_i \qquad [2.1]$$

其中,β_k 是与第 k 个变量($k = 1, \cdots, K$)相关的系数,$Z_i = \sum_{k=1}^{K} \beta_k X_{ik}$。如果 $\beta_k > 0$,对某个特定的个人第 k 个因素取值的增加会导致他/她剥夺分数的上升,反之 $\beta_k < 0$ 会导致分数的下降。但是,因为剥夺分数与导致剥夺的因素之间的关系并不精确——例如,可能有些因素并不包括在方程内或者有些因素的测量并不准确——所以此方程还包括了

一个误差项 ε_i，用以捕捉这种不精确性。

方程 2.1 存在的问题是，用 D_i 的取值所表示的真实的个人剥夺状态的精确图案是很难观察到的。剥夺指数是一个潜在变量，(虽然在概念上有用)但在原则上和实践上都难以察觉，方程 2.1 是一个潜在回归，按照现在的情况来看是无法进行估计的。

但是，我们能观察到的是一个人被剥夺的程度(deprivation level)——例如一个人可以被归类为“没有被剥夺”、“轻度被剥夺”或“严重被剥夺”——而且变量 Y_i 可以以下列方式与这些剥夺程度相关联，$Y_i = 1$ 如果此人没有被剥夺，$Y_i = 2$ 如果此人轻度被剥夺，$Y_i = 3$ 如果此人严重被剥夺。按照前面的讨论，Y_i 是一个有序变量。样本中个人在这三个剥夺程度中的分类内在地基于潜在变量 D_i 的值以及“临界”值 δ_1 和 δ_2，如下所示：

$$\begin{aligned} &Y_i = 1, \text{如果 } D_i \leqslant \delta_1 \\ &Y_i = 2, \text{如果 } \delta_1 \leqslant D_i \leqslant \delta_2 \\ &Y_i = 3, \text{如果 } D_i \geqslant \delta_2 \end{aligned} \qquad [2.2]$$

方程 2.2 中的 δ_1，$\delta_2 \geqslant 0$，是跟方程 2.1 中的 β_k 一起有待估计的未知参数($\delta_1 < \delta_2$)。一个人在剥夺程度上的归类取决于他/她的剥夺分数 D_i 是否跨过某个临界值。Y_i 取值为 1，2 和 3 的概率表示如下：

$$\begin{aligned} &\Pr(Y_i = 1) = \Pr(Z_i + \varepsilon_i \leqslant \delta_1) = \Pr(\varepsilon_i \leqslant \delta_1 - Z_i) \\ &\Pr(Y_i = 2) = \Pr(\delta_1 \leqslant Z_i + \varepsilon_i \leqslant \delta_2) = \Pr(\delta_1 - Z_i < \varepsilon_i \leqslant \delta_2 - Z_i) \\ &\Pr(Y_i = 3) = \Pr(Z_i + \varepsilon_i \geqslant \delta_2) = \Pr(\varepsilon_i \geqslant \delta_2 - Z_i) \end{aligned} \qquad [2.3]$$

N 个观察中的每一次都被视为多项分布中的一次单一

的抽取,在这种情况下,这个多项分布有三种结果,没有被剥夺、轻度被剥夺或严重被剥夺。假设 N 个人,其中 N_1 个人没有被剥夺,N_2 个人轻度被剥夺,N_3 个人严重被剥夺。[6]那么观察到现在这个样本的可能性,就是个体观察值概率的乘积,为:

$$\begin{aligned} L &= [\Pr(Y_i = 1)]^{N_1}[\Pr(Y_i = 2)]^{N_2}[\Pr(Y_i = 3)]^{N_3} \\ &= [F(\delta_1 - Z_i)]^{N_1}[F(\delta_2 - Z_i) - F(\delta_1 - Z_i)]^{N_2} \\ &\quad \times [1 - F(\delta_1 - Z_i)]^{N_3} \end{aligned} \qquad [2.4]$$

其中,$F(x) = \Pr(\varepsilon_i < x)$ 是误差项的累积概率分布。如果我们已知误差项的概率分布——即我们知道什么是 $F(x)$——那么我们就可以像 β_k, δ_1,和 δ_2 的估计值一样,选择那些使观察到样本结果的可能性最大化的值[7]。若缺乏这样的信息,我们可以假定误差项服从一个特定的概率分布。

次序 logit 与次序 probit 模型之间的差别在于:方程 2.1 中的误差项 ε_i 的(假定的)分布不同。次序 logit 模型是假定 ε_i 具逻辑分布,而次序 probit 模型则假定 ε_i 具正态分布。很自然地,我们会问哪种分布更适合。[8]逻辑分布除了在末尾部分比正态分布大很多之外,其他与正态分布相似。[9]正如格林(2000:815)指出的:“要在理论上证明选择一种而非另一种分布更为合理是很困难的事……在大部分应用中,(它们)看起来并没有什么大的差别。”

使用系数 β_k 的估计值 $\hat{\beta}_k$ 可以让我们对样本中的每个个体计算估计值 $\hat{Z}_i = \sum_{k=1}^{k} \hat{\beta}_k X_{ik}$。使用 $\hat{Z}_i$ 以及临界参数 δ_1, δ_2 的估计值 $\hat{\delta}_1$ 和 $\hat{\delta}_2$,可以对样本中每个个体处于不同剥夺程度的概率进行估计。这些估计可分别用 $\hat{P}_{i1}$, $\hat{P}_{i2}$ 和 $\hat{P}_{i3}$

表示，计算如下：

$$\hat{P}_{i1} = \Pr(\varepsilon_i \leqslant \hat{\delta}_1 - \hat{Z}_i) = F(\hat{\delta}_1 - \hat{Z}_i) \qquad [2.5a]$$

$$\hat{P}_{i2} = \Pr(\hat{\delta}_1 - \hat{Z}_i < \varepsilon_i \leqslant \hat{\delta}_2 - \hat{Z}_i)$$
$$= F(\hat{\delta}_2 - \hat{Z}_i) - F(\hat{\delta}_1 - \hat{Z}_i) \qquad [2.5b]$$

$$\hat{P}_{i3} = \Pr(\varepsilon_i \geqslant \hat{\delta}_2 - \hat{Z}_i) = 1 - F(\hat{\delta}_2 - \hat{Z}_i) \qquad [2.5c]$$

其中，对每个 $i = 1, \cdots, N$, $\sum_{j=1}^{3} \hat{p}_{ij} = 1$。

上面描述的模型同时也称做比例比数模型(proportional-odds model)，因为如果我们考虑某个类别 $j = m$ 的比数比(OR)，$OR(m) = \frac{\Pr(Y_i \leqslant m)}{\Pr(Y_i > m)}$，那么这个比数比就独立于类别 m，即不随着 m 的变化而变化。比数比被假定为对所有的类别保持不变[10]。

关于符号的说明

用 STATA 估计的次序回归(不管是 logit 还是 probit)并不明显包含截距项：或者说，方程 2.1 中的 $\beta_k(k = 1, \cdots, K)$ 全部都是斜率。截距项没有明显地表示出来，因为如下所示它被并入到临界值 δ_1 和 δ_2 中了。另一方面，格林(2000：876)在用公式表示次序回归时，明显地包括了截距项。对应于方程2.1，他的方程是：

$$D_i = \beta_0 + \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i = \beta_0 + Z_i + \varepsilon_i = W_i + \varepsilon_i \qquad [2.6]$$

其中 β_0 是截距项，$W_i = \beta_0 + Z_i$。格林的临界值用 μ_1 和

μ_2 表示(跟 STATA 里面的临界值 δ_1 和 δ_2 不一样,究竟如何不一样表现如下)。格林(2000)设第一个临界值 μ_1 等于0。[11]所以在他的表述中,表示概率 $\hat{P}_{i1}$, $\hat{P}_{i2}$ 和 $\hat{P}_{i3}$ 的方程为:

$$\hat{P}_{i1} = \Pr(\varepsilon_i \leqslant -\hat{W}_i) = F(-\hat{W}_i) = F(-\hat{\beta}_0 - \hat{Z}_i) = F(\hat{\delta}_1 - \hat{Z}_i) \quad [2.7a]$$

$$\hat{P}_{i2} = \Pr(-\hat{W}_i < \varepsilon_i \leqslant \hat{\mu}_2 - \hat{W}_i) = F(\hat{\mu}_2 - \hat{W}_i) - F(-\hat{W}_i) = F(\hat{\mu}_2 - \hat{\beta}_0 - \hat{Z}_i) - F(-\hat{\beta}_0 - \hat{Z}_i) = F(\hat{\delta}_2 - \hat{Z}_i) - F(\hat{\delta}_1 - \hat{Z}_i) \quad [2.7b]$$

$$\hat{P}_{i3} = \Pr(\varepsilon_i \geqslant \hat{\mu}_2 - \hat{W}_i) = 1 - F(\hat{\mu}_2 - \hat{W}_i) = 1 - F(\hat{\mu}_2 - \hat{\beta}_0 - \hat{Z}_i) = 1 - F(\hat{\delta}_2 - \hat{Z}_i) \quad [2.7c]$$

当 $\delta_i = \mu_i - \beta_0$ 时,STATA 方程 2.5a 至 2.5c 与格林(2000)的方程 2.7a 至 2.7c 是相同的。[12]也就是说,STATA 的临界值等于格林的临界值减去截距项。从这种意义上说,STATA 把截距项合并到临界值中去了。[13]

次序 logit

在逻辑分布的情况下,随机变量 X 的累积分布函数是:

$$\Pr(X \leqslant x) = \Lambda(x) = \exp(x)/[1 + \exp(x)] = 1/(1 + \exp(-x)) \quad [2.8]$$

如果假设误差项服从逻辑分布,

$$\Pr(Y_i = 1) = \Lambda(\delta_1 - Z_i) = 1/[1 + \exp(Z_i - \delta_1)] \quad [2.9a]$$

$$\Pr(Y_i = 2) = \Lambda(\delta_2 - Z_i) - \Lambda(\delta_1 - Z_i)$$

$$= 1/[1+\exp(Z_i-\delta_2)]-1/[1+\exp(Z_i-\delta_1)] \quad [2.9b]$$

$$\Pr(Y_i=3)=1-\Lambda(\delta_2-Z_i)=1-1/[1+\exp(Z_i-\delta_2)] \quad [2.9c]$$

则通过使用逻辑分布函数 $\Lambda(.)$代替 $F(.)$最大化似然函数(方程 2.4)，获得 β_k，δ_1 和 δ_2 的估计值。

次序 probit

标准正态变量[14](SNV)X 的累积分布是：

$$\Pr(X<x)=\Phi(x)=\int_0^x(1/2\pi)\exp(-X^2/2)dX \quad [2.10]$$

如果假设误差项是 SNVs，

$$\Pr(Y_i=1)=\Phi(\delta_1-Z_i) \quad [2.11a]$$

$$\Pr(Y_i=2)=\Phi(\delta_2-Z_i)-\Phi(\delta_1-Z_i) \quad [2.11b]$$

$$\Pr(Y_i=3)=1-\Phi(\delta_2-Z_i) \quad [2.11c]$$

则对 β_k，δ_1 和 δ_2 的估计通过使用正态分布函数 $\Phi(.)$代替 $F(.)$最大化似然函数(方程 2.4)获得。

边际效应:连续变量

一个意料之中的问题是当其中一个影响结果的变量的值改变的时候，出现不同结果的概率是怎么变化的。例如，如果年纪是影响剥夺的一个因素，那么个人处于不同的剥夺

程度(没有被剥夺、轻度被剥夺或严重被剥夺)的概率是怎么受他/她是大一岁还是小一岁影响的? 在逻辑分布下,X_{ik}(第 i 个人的第 k 个决定变量的值)的一个小变化对个人 i 的这三种概率的边际效应是:

$$\frac{\partial \Pr(Y_i=1)}{\partial X_{ik}} = \frac{d}{dZ_i}[\Lambda(\delta_1 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = -\Lambda'(\delta_1 - Z_i)\beta_k \quad [2.12a]$$

$$\frac{\partial \Pr(Y_i=2)}{\partial X_{ik}} = \frac{d}{dZ_i}[\Lambda(\delta_2 - Z_i) - \Lambda(\delta_1 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = [\Lambda'(\delta_2 - Z_i) - \Lambda'(\delta_1 - Z_i)]\beta_k \quad [2.12b]$$

$$\frac{\partial \Pr(Y_i=3)}{\partial X_{ik}} = \frac{d}{dZ_i}[1 - \Lambda(\delta_2 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = \Lambda'(\delta_2 - Z_i)\beta_k \quad [2.12c]$$

在正态分布下,则为:

$$\frac{\partial \Pr(Y_i=1)}{\partial X_{ik}} = \frac{d}{dZ_i}[\Phi(\delta_1 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = -\Phi'(\delta_1 - Z_i)\beta_k \quad [2.13a]$$

$$\frac{\partial \Pr(Y_i=2)}{\partial X_{ik}} = \frac{d}{dZ_i}[\Phi(\delta_2 - Z_i) - \Phi(\delta_1 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = [\Phi'(\delta_2 - Z_i) - \Phi'(\delta_1 - Z_i)]\beta_k \quad [2.13b]$$

$$\frac{\partial \Pr(Y_i=3)}{\partial X_{ik}} = \frac{d}{dZ_i}[1 - \Phi(\delta_2 - Z_i)]\frac{\partial Z_i}{\partial X_{ik}} = \Phi'(\delta_2 - Z_i)\beta_k \quad [2.13c]$$

其中 $\Lambda'(x) = d\Lambda(x)/dx$ 和 $\Phi'(x) = d\Phi(x)/dx$ 分别是逻辑分布和正态分布的概率密度函数。边际效应可以通过估计在相关点上的合适的密度函数并乘以相关的系数获得。举例,由方程 2.8,逻辑分布的密度函数是:

$$\Lambda'(x)=\frac{d}{dx}\left[\frac{\exp(x)}{1+\exp(x)}\right]=\frac{[1+\exp(x)]\exp(x)-[\exp(x)]^2}{[1+\exp(x)]^2}$$
$$=\Lambda(x)[1-\Lambda(x)]$$

方程 2.12a 的边际效应是：

$$\Lambda'(\delta_1-Z_i)\beta_k=\Lambda(\delta_1-Z_i)[1-\Lambda(\delta_1-Z_i)]\beta_k$$
$$=\frac{1}{1+\exp(Z_i-\delta_1)}\left[1-\frac{1}{1+\exp(Z_i-\delta_1)}\right]\beta_k$$

现在如果第 k 个决定变量的值增加一点且 $\beta_k>0$，那么在 logit 和 probit 模型下，没有被剥夺的概率都一定会降低，因为通过方程 2.12a 和 2.13a，$\Pr(Y_i=1)$ 的推导与 β_k 方向相反。同时在两个模型中，严重被剥夺时概率都会上升，因为通过 2.12c 和 2.13c，$\Pr(Y_i=3)$的推导与 β_k 方向相同。但是我们并不清楚中间那个概率会发生什么变化。取决于其他两个概率怎么变化，轻度被剥夺的概率，$\Pr(Y_i=2)$，可能上升、下降或者保持不变。[15]因此，给定某个决定变量的变化，不可能根据相关系数的方向对所有概率的变化方向做出推断。只有处于两个极端情形的概率的变化方向能被确定。由于这个原因，格林(2000:878)警告说，“我们在解释这个模型的系数时必须非常小心，因为这是模型中最不明显的”。

边际效应:虚拟变量

然而，上述估计边际效应的方法只在决定变量为连续变量而非虚拟变量时适用。虚拟变量的作用应该通过比较虚拟变量取其中一个值时结果出现的概率和它取另外一个值的结果出现的概率进行分析，且这种比较要在其他变量的值

保持不变的情况下进行。例如,假设雇佣状态是影响剥夺的一个因素,$X_{ik}=1$ 表示一个人失业,$X_{ik}=0$ 则表示他/她受雇。为了分析如果他/她从受雇转为失业如何影响其处于三种不同的剥夺程度的概率,首先使用方程 2.9a 至 2.9c,假设 $X_{ik}=1$ 对 Z_i 进行估计(称为 Z_i^1);如果应用 probit 模型,则使用方程2.11a至 2.11c,计算 $Y_i=1$(没有被剥夺),$Y_i=2$(轻度被剥夺)和 $Y_i=3$(严重被剥夺)这三种概率。然后,保持其他决定变量的值不变,假设 $X_{ik}=0$ 对 Z_i 进行估计(称为 Z_i^0),重新计算以上三种概率。注意,由方程 2.1,$Z_i^1=Z_i^0+\beta_k$。这两组概率之间的区别就是当个人从受雇状态($X_{ik}=0$)转为失业状态($X_{ik}=1$),或从失业转为受雇,对他/她处于不同剥夺程度的概率的作用。

平行斜率假设

次序 logit 和 probit 模型的一个关键假设是方程 2.1 的斜率系数 β_k 并不随着所关注的剥夺结果的变化而改变。也就是说,次序 logit 和 probit 模型符合一种平行斜率累积模型,这种模型在 logit 的情况下,采取下列形式:

$$\log\frac{p_1}{1-p_1}=\alpha_1+\sum_{k=1}^{K}\beta_k X_{ik}$$

$$\log\frac{p_1+p_2}{1-p_1-p_2}=\alpha_2+\sum_{k=1}^{K}\beta_k X_{ik}$$

$$\cdots\cdots$$

$$\log\frac{p_1+p_2+\cdots+p_{M-1}}{1-p_1-p_2-\cdots-p_{M-1}}=\alpha_M+\sum_{k=1}^{K}\beta_k X_{ik}$$

$$p_1+p_2+\cdots+p_M=1$$

其中，$p_j = \Pr(Y_i = j)$。通过对数据进行多类别 logit 模型估计，我们可以对平行斜率假设的有效性进行检验（见第3章）。多类别 logit 模型允许斜率系数 β_k 在不同的结果 $j = 1, \cdots, M$ 之间不同。当次序 logit 模型估计 K 个系数时，多类型 logit 模型估计 $K(M-1)$ 个系数。[16] 如果 L_1 是从次序 logit 模型得出的概率值，L_2 是从多类别 logit 模型得出的概率值，那么我们就能计算 $2(L_2 - L_1)$，并与 $\chi^2(K(M-1))$ 进行比较。注意，这并不是严格的似然比检验，因为次序 logit 模型并不嵌套在多类别 logit 模型之中。因此，这个检验只是提示性的：一个“非常大的”χ^2 值是引起担心的依据，若 χ^2 值不是很大则不用担心（STATA，1990：480）。但是，如果我们确实有理由相信平行斜率假设并不有效，那么此模型就该用多类别 logit 的方法进行估计，就算事实上因变量是明显有序的。

第3节 | 应用:剥夺状态

界定与构造剥夺指数

有 N 个人,标号为 $i = 1, \cdots, N$。如果某种条件的出现导致一个人经历剥夺,那么,这种条件就定义为"剥夺导致条件"(DIC)。假设有 K 种 DIC,标号为 $k = 1, \cdots, K$,I_{ik} 是关于 DIC k 及个人 i 的一个类别变量,$I_{ik} = 1$ 表示 DIC 对第 i 个人出现了,$I_{ik} = 0$ 表示没有出现。那么对个人 i 的剥夺程度,记为 D_i^*,定义为:

$$D_i^* = \sum_{k=1}^{K} \alpha_k^* I_{ik} \qquad [2.14]$$

其中 $\alpha_k^* > 0$ 是赋予第 k 个 DIC 的权重,且独立于所考虑的个人。如果与个人 DIC 相关的权重定义为 $\alpha_k^* = 1 - p_k$,其中 p_k 代表条件 k 发生的频数,那么 α_k^* 包含了"相对剥夺"的概念,即一个特定的 DIC 被经历的频数越小,那么当它被经历的时候赋予它的权重就越大。这些权重的使用响应了德赛和沙(Desai & Shah, 1988)的工作,他们在重新检验汤森(Townsend, 1979)的原始数据时,从根本上主张被剥夺了几乎所有人拥有的东西要比被剥夺了很少人拥有的东西影响更大。

只在偶然的情况下，权重 α_k^* 总和为 1(sum to unity)。但是，它们却可通过定义 $\alpha_k = \alpha_k^* / \Omega$ 标准化，其中 $\Omega = \sum_{k=1}^{K} \alpha_k^*$。在这种标准化之下，个人 i 的剥夺程度可以定义为：

$$D_i = D_i^* \Omega^{-1} = \left(\sum_{k=1}^{K} \alpha_k^* I_{ik}\right)\Omega^{-1} = \sum_{k=1}^{K} \alpha_k I_{ik} \quad [2.15]$$

因为 D_i 只是 D_i^* 的一种简单的标量转换，所以可以通过使用 D_i，与使用 D_i^* 一样，获得个人在他们的剥夺程度上的相同排序。但是，在标准化权重方面，剥夺指数 D_i 比 D_i^* 提供了更好的解释，所以随后的分析将会采用 D_i 这种标准。因为，在定义上 $\sum_{k=1}^{K} \alpha_k = 1$，所以方程 2.15 暗示着 $0 \leqslant D_i \leqslant 1$：当 $D_i = 0$ 时，表示没有任何一种 DIC 出现，也就是当对所有的 $k = 1, \cdots, K$，$I_{ik} = 0$ 时；当 $D_i = 1$ 时，表示所有的 DIC 都出现，也就是当对所有的 $k = 1, \cdots, K$，$I_{ik} = 1$ 时。

构造剥夺指数的一个主要问题在于决定应该加入哪些 DIC。回顾关于构造剥夺指数的文献，诺兰和惠兰(Nolan & Whelan, 1996)指出，鉴赏力的作用是一个主要的问题。如果在生活模式中观察到的差异可以大部分归因于偏好而非资源，那么在构造剥夺指数时，若某些条目不被包含在内，也不能说明是因为需求的原因。例如，皮亚乔德(Piachaud, 1987)强调了处于相似收入水平的家庭在剥夺分数间存在的巨大差异。即使我们可以把偏好从需求中分离出来，研究者在选择把哪些条目放入到剥夺指数中时所起的作用又是另一个更深层次的问题。这样，问题就在于剥夺的测量是应该

单单基于个人自己的环境,还是应根据他/她的更宽泛的社会以及地理环境。例如,布鲁雅和卡克(Borooah & Carcach, 1997)注意到居住社区质量在决定人们对犯罪的害怕程度上的重要性。最后,如果我们可以取得一系列令人满意的指标,还有一个问题在于如何在构造综合指数时对这些指标进行加权。数据本身的限制在这些问题上起中心作用;我们只能从现成可用的数据中构建剥夺指数,而不能从我们希望可得到的数据中构建该指数。作者所构建的剥夺指数基于1991 年人口普查①的数据,该数据包含了在此地居住的13164 个人的生活环境的信息。[17] 比如,这些数据告诉我们在普查时个人是否:

住在一个通常没有任何成员有机会使用一辆小车或小巴的家庭中;

住在一个没有任何成员是挣钱者的家庭中;

住在一个家庭成员没有专用的室内厕所的家庭中;

住在一个没有中央暖气系统的房子里;

住在一个没有室内公共供水的房子里;

住在一个没有与公共排水管道连通的房子里;

住在一个代表非永久住处的房子里;

住在一个居住人数与房间数量的比例大于 1 的房子里;

长年患有疾病,面临健康问题或者残疾,这些问题限制了他/她的日常活动或者可能做的工作。

---

① 北爱尔兰 1991 年的人口普查。——译者注

利用这些信息，对退休者和非退休者分别构建了剥夺指数。这两组指数都围绕着相同的一组 DIC 构建，但是对这些 DIC 赋予的权重随着个人是否退休而不同。正如上面提到的，与某个 DIC 相关的权重反映了相关 DIC 被经历的频数——频数越低，权重越大。但是，其中一些 DIC 被经历的频数在已退休样本和非退休样本之间存在很大差异，在对退休者和非退休者构造剥夺指数时使用不同的权重反映了这种差异。取决于对他/她计算出来的剥夺指数的值，每个人都被指定了三个剥夺程度——没有被剥夺、轻度被剥夺、严重被剥夺——中的一种，[18] 与每个结果($j=1, 2, 3$)和每个人($i=1, \cdots, N$)相关的是有序因变量 Y_i 的如下取值：

$Y_i = 1$ 如果这个人没有被剥夺

$Y_i = 2$ 如果这个人轻度被剥夺

$Y_i = 3$ 如果这个人严重被剥夺

应该特别强调的是，正是这些关于 Y_i 的值的数据典型地可供研究者分析利用。在 13164 个人的全部样本中，45.9%(6042 个人)被归为没有被剥夺，34.9%(4594 个人)被归为轻度被剥夺，19.2%(2528)被归为严重被剥夺。

方程设定

用于“解释”个人的剥夺程度的决定变量为：

AGE_i，以年为单位，通过设定 16 岁的人 $AGE_i=0$ 进行

标准化;[19]

$HIGHED_i = 1$,如果这个人具有英国标准的学士或者更高的学位资格;否则 $HIGHED_i = 0$;

$MIDED_i = 1$,如果这个人有高于 A 程度文凭(post-A level)①,但低于学位文凭;[20]否则 $MIDED_i = 0$;

$RET_i = 1$,如果这个人退休了,否则 $RET_i = 0$;

$INAC_i = 1$,如果这个人在经济上不活跃②,否则 $INAC_i = 0$;

$UE_i = 1$,如果这个人失业,否则 $UE_i = 0$;

$HNUM_i = 1$,如果这个家庭的人数大于或等于 6,否则 $HNUM_i = 0$;

$SNPAR_i = 1$,如果这个人是单身父亲/母亲,否则 $SNPAR_i = 0$;

$AREA_i = 1$,如果这个人居住在北爱尔兰的 a 地区,否则 $AREA_i = 0$;在北爱尔兰 1991 年的普查中一共有 19 个这样的地区。[21]

除了这些变量外,剥夺程度还可能取决于个人的环境,因为在北爱尔兰,天主教徒是一个相对来说处于劣势的群体,所以也还可能取决于他/她的宗教信仰(在样本的 13164 个人中,7243 个是男性,5921 个是女性,4364 个是天主教徒,8889 个是新教徒[22])。为了说明这点,还应考虑另外两个变量:

---

① 英国文凭高级水平(A-level)是中学和大学的衔接课程,其考试成绩是从中学升人大学的考核标准;在这个文凭和学位文凭之间,还有大学预科文凭(International Foundation Diploma, IFD)——部分大学设有该课程,以及高级国家文凭(Higher National Diploma, HND)——相当于中国的大专文凭。——译者注

② 即不从事经济活动。——译者注

$SEX_i = 1$，如果这个人是女性，否则 $SEX_i = 0$；

$CT_i = 1$，如果这个人是天主教徒，否则 $CT_i = 0$。

因此，在这个应用情况下，方程 2.1 设定为：

$$\begin{aligned} D_i = {} & \beta_1 + \beta_2 * SEX_i + \beta_3 * CT_i + \beta_4 * AGE_i + \beta_5 * AGE_i^2 \\ & + \beta_6 * HIGHED_i + \beta_7 * MIDED_i + \beta_8 * RET_i \\ & + \beta_9 * INAC_i + \beta_{10} * UE_i + \beta_{11} * HNUM_i \\ & + \beta_{12} * SNPAR_i + \beta_{13} * AREA_{2i} \cdots \\ & + \beta_{21} * AREA_{10i} + \varepsilon_i = Z_i + \varepsilon_i \end{aligned} \qquad [2.16]$$

年龄变量的平方（上面的 AGE_i^2）引入了年龄的非线性作用：年龄的增加对 D_i 的边际作用取决于这个增加是在哪个年龄发生的。如果 $\beta_4 < 0$ 且 $\beta_5 > 0$，那么一个人年龄的增加降低了他/她的剥夺分数，但是这个人年纪越大这种影响就越小。[23]

方程统计原理

估计参数 $\hat{\beta}_k$、$\hat{\delta}_1$ 和 $\hat{\delta}_2$ 使观察到样本的可能性最大化，其中样本 $N_1 = 6042$，$N_2 = 4594$，$N_3 = 2528$（见方程 2.4）。

这些估计值列于对应次序 logit 模型的表 2.1 以及对应次序 probit 模型的表 2.2 中。表 2.1 和表 2.2 中的 z-ratio是估计系数与他们的估计标准误差之间的比：这些 z-ratio（渐近）服从零假设为相关系数等于零的 $N(0, 1)$ 分布。[24]

表 2.1 北爱尔兰社会剥夺程度的次序 logit 模型结果

Ordered Logit Estimates; *Log Likelihood* = −12423.562; *Number of obs* = 13164; *LR* $\chi^2(20)$ = 2571.98; *Prob*$>\chi^2$ = 0.0000; *Pseudo* R^2 = 0.0938

y	*Coefficient*	*Standard Error*	*z*	$P>\|z\|$	[95% *Conf.*	*Interval*]
sex	−0.1649064	0.0360722	−4.572	0.000	−0.2356066	−0.0942061
ct	0.1829855	0.0393496	4.650	0.000	0.1058617	0.2601092
age	−0.0253924	0.0049969	−5.082	0.000	−0.0351862	−0.0155987
age2	0.0008173	0.0000946	8.638	0.000	0.0006319	0.0010028
ret	0.6264438	0.0875236	7.157	0.000	0.4549007	0.797987
inac	1.297307	0.0765827	16.940	0.000	1.147207	1.447406
ue	1.460677	0.0619881	23.564	0.000	1.339182	1.582171
highed	−0.9081216	0.0683311	−13.290	0.000	−1.042048	−0.774195
mided	−0.6692574	0.1230722	−5.438	0.000	−0.9104745	−0.4280403
hnum	0.9214472	0.0551311	16.714	0.000	0.8133923	1.029502
snpar	0.305537	0.0637575	4.792	0.000	0.1805746	0.4304994
ard	−0.4497501	0.0642359	−7.002	0.000	−0.5756502	−0.3238499
dwn	0.0492598	0.0680278	0.724	0.469	−0.0840723	0.1825919
crk	−0.3161319	0.0706348	−4.476	0.000	−0.4545735	−0.1776903
ant	0.3472878	0.0717874	4.838	0.000	0.206587	0.4879887
col	0.4384781	0.0715933	6.125	0.000	0.2981578	0.5787985
arm	0.3824736	0.0740609	5.164	0.000	0.237317	0.5276303
ban	0.2364069	0.066677	3.546	0.000	0.1057223	0.3670914
dry	0.0791624	0.0787678	1.005	0.315	−0.0752198	0.2335445
frm	0.6350174	0.0747171	8.499	0.000	0.4885745	0.7814602
_cut1	0.1804476	0.074681		(Ancillary parameters)		
_cut2	2.058464	0.0772157				

y	*Probability*	*Observed*
1	Pr($xb+u<$ _cut1)	0.4591
2	Pr(_cut1 $<xb+u<$ _cut2)	0.3493
3	Pr(_cut2 $<xb+u$)	0.1917

表 2.2 北爱尔兰社会剥夺程度的次序 Probit 模型结果

Ordered Probit Estimates; *Log Likelihood* = −12426.255; *Number of obs* = 13164;
LR $\chi^2(20)$ = 2566.60; *Prob* > χ^2 = 0.0000; *Pseudo* R^2 = 0.0936

y	*Coefficient*	*Standard Error*	*z*	*P* > |*z*|	[95% *Conf.*	*Interval*]
sex	−0.1003127	0.021542	−4.657	0.000	−0.1425343	−0.0580911
ct	0.112619	0.0235271	4.787	0.000	0.0665067	0.1587314
age	−0.0154206	0.002965	−5.201	0.000	−0.021232	−0.0096093
age2	0.0004965	0.000056	8.872	0.000	0.0003868	0.0006062
ret	0.3679154	0.0528714	6.959	0.000	0.2642894	0.4715415
inac	0.7680422	0.0445939	17.223	0.000	0.6806398	0.8554446
ue	0.8555121	0.0357888	23.904	0.000	0.7853673	0.925657
highed	−0.5562051	0.0401161	−13.865	0.000	−0.6348313	−0.4775789
mided	−0.4048568	0.0721096	−5.614	0.000	−0.546189	−0.2635246
hnum	0.5459127	0.0328011	16.643	0.000	0.4816237	0.6102017
snpar	0.1748777	0.0379915	4.603	0.000	0.1004157	0.2493396
ard	−0.2729771	0.03806	7.172	0.000	−0.3475734	−0.1983808
dwn	0.0216987	0.0403927	0.537	0.591	−0.0574696	0.100867
crk	−0.1831732	0.0417875	4.383	0.000	−0.2650752	−0.1012712
ant	0.2026288	0.0428712	4.726	0.000	0.1186028	0.2866547
col	0.2518903	0.0428189	5.883	0.000	0.1679668	0.3358137
arm	0.2122576	0.0444114	4.779	0.000	0.1252129	0.2993024
ban	0.129325	0.0399411	3.238	0.001	0.0510418	0.2076082
dry	0.0360175	0.0471155	0.764	0.445	−0.0563271	0.1283622
frm	0.3626456	0.0447755	8.099	0.000	0.2748871	0.450404
_cut1	0.1009536	0.0444785		(Ancillary parameters)		
_cut2	1.212453	0.0453671				

y	*Probability*	*Observed*
1	Pr(*xb* + *u* < _cut1)	0.4591
2	Pr(_cut1 < *xb* + *u* < _cut2)	0.3493
3	Pr(_cut2 < *xb* + *u*)	0.1917

格林(2000:831—833)提供了一些测量离散因变量方程的"拟合优度"的建议。他提议，至少我们应该报告对数似然函数最大化了的值。L_1 的值就是最大对数似然值，列于表格的顶部(分别为−12423.56 和−12426.25)。由于方程中所有斜率都为

零的假设经常令人很感兴趣,所以还应该报告"完整"模型和"只有截距项"模型的比较结果。表 2.1 和表 2.2 顶部的 χ^2 值(分别为 2571.98 和 2566.6)等于 $2(L_1 - L_0)$,其中 L_0 是当解释变量只有常数项时的对数似然函数的值,L_1 则如前面观察到的,是当所有的解释变量都包括在内时对数似然函数的值。自由度等于被估计的斜率系数的个数。这些 χ^2 的值完全拒绝了这个模型不比"只有截距项"的模型更具解释力的零假设。[25]

"pseudo-R^2"被定义为 $1 - L_1/L_0$,这归功于麦克法登(McFadden, 1973)。这个值的范围固定在 0 与 1 之间。值为 0 对应于所有的斜率系数均为 0,值为 1 对应于理想预测(也就是,$L_1 = 0$)。遗憾的是,正如格林(2000)指出的,介于 0 和 1 之间的值并没有自然的解释,虽然提示是 pseudo-R^2 的值随着模型拟合的改善而增加。还有人提出其他方法。本阿齐瓦和勒曼(Ben-Akiva & Lerman, 1985)以及凯和利特尔(Kay & Little, 1986)提出了一种通过预测规则测量正确预测的平均概率的拟合方法。卡默(Carmer, 1999)提出了一种纠正本阿齐瓦/勒曼测量不足之处的方法,考虑了在不平衡样本中,对频数少的结果进行的预测往往非常糟糕这样一个事实。维尔和齐默尔曼(Veall & Zimmerman, 1996)在他们的 pseudo-R^2 测量中,认为在多类别 probit 或者 logit 类型的模型中,只有麦克法登(1973)的方法"看起来有价值"。

另外一种拟合优度的"点"测量方法可能是评估模型的预测能力。这些评估在二分选择模型中是例行程序,把按某种规则(如当 $\hat{P}_i > 0.5$ 时 $Y_i = 1$,否则 $Y_i = 0$)预测到的命中($Y_i = 1$)和不命中($Y_i = 0$)与真实的命中与不命中进行比较。这个方法可以扩展到多元结果模型,其预测可以基于

如果 $\hat{P}_{im} = \text{Max}_j(\hat{P}_{ij})$，则 $Y_i = m$ 的规则。然后这些预测可以与估计所有情况都处于因变量最常见类别中的这样一种“简单”模型进行比较，从这个简单模型到完整模型减少了的误差的百分比就可以计算出来了。但是，这种方法并不是没有缺陷。首先，与线性回归模型中系数被选择以最大化 R^2 的情况不同，在离散选择模型中，系数的估计值并不最大化任何拟合优度测量。所以，不管拟合优度怎么被测量，以拟合优度为基础对模型进行评估都可能是误导性的。第二，这些预测严重取决于采取的预测规则，但可能这些预测规则本来非常不适合样本的需要。比如，在一个二分模型中，如果样本是不平衡的——就是说有更多的 1，更少的 0——则模型应该预测估计概率最大的结果这样的一个规则就可能永远无法预测到一个 1(或者 0)。

估计值

表 2.1 和表 2.2 中的所有系数，除了两个与居住在 County Down 和 County Derry 地区(表中的 *dwn* 和 *dry* 变量)相关的系数之外，其他都明显不等于 0。在表 2.1 的“$P > |z|$”这列下面的数字表示，在 *dwn* 和 *dry* 的系数等于 0 的零假设下，分别有 46.6%和 31.5%的机会，会观察到超出观测值 0.0492598 和 0.0791624 的值。在表 2.2 的“$P > |z|$”这列下面的数字表示，在这两个系数等于 0 的零假设下，观察到一个超出观测值的数值的机会为“零”。在“95% Conf. Interval”这列下面的数字表示如果它们的相关系数要在 5%显著性水平下不等于零，估计值必须超出的上下限。在表 2.1 和表 2.1 中，对 *dwn* 和 *dry* 系数的估计值处于它们的 95%区

间内,而其他系数的估计值却在它们的区间之外。

在解释单个系数的时候,必须强调其他条件不变("其他事情相同")的重要性。变量 *ct* 的估计值是正的,表示在其他条件不变的情况下,天主教徒与新教徒相比,有更高的被严重剥夺的概率和更低的没有被剥夺的概率。这并不是说每个天主教徒都比每个新教徒有更高的被严重剥夺的概率。相反,正确的说法应该是,给定两个除了在宗教信仰之外其他每种特征都相似的人,天主教徒比新教徒更加可能被严重剥夺,更加不可能没有被剥夺。变量 *sex* 的估计值是负的,表示其他条件相同的女性比男性有更低的被严重剥夺的概率和更高的没有被剥夺的概率。其他系数的解释相似。比如,变量 *highed* 和 *mided* 的负的估计值表示,在其他条件相同的情况下,具有高学历文凭的人比没有高学历文凭的人有更低的被严重剥夺的概率和更高的没有被剥夺的概率。正如前面的讨论所说明的,系数估计值的方向只允许对处于两端的结果,其概率随着相关变量值的改变而变化的方向进行估计。处于中间的结果其概率的变化方向不能推断。比如,从对估计值的审查中,我们无法判断女性相对于男性来说被轻度剥夺的概率更大还是更小。

临界点

估计出来的临界值呈现于表中各估计值的下面。表 2.1 和表 2.2 的下方显示的部分表示各类别的概率如何从拟合方程中计算出来。[26] 这部分显示的结果通过在 oprobit 和 ologit 命令语法中加入"table"这个选择命令而得,这些命令列于第 4 章的程序表中。有序回归背后的假设是观察到的

类别代表了粗略的但以一种隐含的等级正确排序的差别。属于这些类别的概率基于潜在变量的值跨越特定界限的概率，这些界限是通过临界点的值而建立的。

在次序 logit 模型中，这两个临界点被估计为[27] $_cut1 = 0.1804476$ 和 $_cut2 = 2.058464$，在方程 2.5a 至方程 2.5c 的符号中，$\hat{\delta}_1 = _cut1$，$\hat{\delta}_2 = _cut2$。因此，在次序 logit 模型中，一个人没有被剥夺的概率是 $\Pr(\hat{Z}_i + \varepsilon_i \leqslant 0.1804476)$，被轻度剥夺的概率为 $\Pr(0.1804476 \leqslant \hat{Z}_i + \varepsilon_i \leqslant 2.058464)$，被严重剥夺的概率为 $\Pr(\hat{Z}_i + \varepsilon_i \geqslant 2.058464)$。在次序 probit 模型中，这两个临界点被估计为 $_cut1 = 0.1009536$ 和 $_cut2 = 1.212453$。在这个模型中，一个人没有被剥夺的概率是 $\Pr(\hat{Z}_i + \varepsilon_i \leqslant 0.1009536)$，被轻度剥夺的概率为 $0.1009536 \leqslant \hat{Z}_i + \varepsilon_i \leqslant 1.212453)$，被严重剥夺的概率为 $\Pr(\hat{Z}_i + \varepsilon_i \geqslant 1.212\ 453)$。这两个模型中的临界点不一样的原因是斜率系数的估计值也完全不同。但是，每个模型中估计出来的斜率和临界系数的结合使这两个模型得出的预测非常相似。

概率估计值：个人的估计

用估计到的 $\hat{Z}_i$——记住 $\hat{Z}_i = \sum_{k=1}^{k} \hat{\beta}_k X_{ik}$，可利用表 2.1 中的估计值，结合每个人的决定变量的取值而得出——STATA 会对样本 13164 个人中的每个人，通过分别计算方程 2.5a，2.5b 和 2.5c 中的 $\hat{p}_{i1}$，$\hat{p}_{i2}$ 和 $\hat{p}_{i3}$，对他们处于三种不同的剥夺程度的概率进行预测。表 2.3 显示了通过误差项 ε_i 的逻辑分布对样本前 25 个人的估计结果。从表 2.3 我们可以看到，给定他/她的情况，第 17 个人有非常低的被严重剥夺的概率(4%)和非常高的没有被剥夺的概率(78%)。另

一方面,第二个人的情况表示他/她有非常高的被严重剥夺的概率(69%)和非常低的没有被剥夺的概率(6%)。把 logit 模型预测的概率与假设 ε_i 为正态分布而得出的概率进行比较是非常有启发性的。表 2.4 列出了后者的估计结果。对表 2.3 和表 2.4 进行的比较显示,尽管系数估计值在 logit 和 probit 模型之间存在差异,预测到的概率却非常相似。

表 2.3 处于不同剥夺水平的次序 Logit 概率计算值:样本前 25 个观测值

P_{num}	$p1$	$p2$	$p3$
1	0.2468799	0.4350583	0.3180617
2	0.0635542	0.2438714	0.6925744
3	0.1456977	0.3815914	0.472711
4	0.221723	0.4290327	0.3492444
5	0.1984064	0.4197503	0.3818433
6	0.0774309	0.2769676	0.6456016
7	0.3072922	0.4363922	0.2563156
8	0.6020886	0.3061397	0.0917717
9	0.6067917	0.303063	0.0901453
10	0.5668221	0.3285579	0.10462
11	0.3003936	0.4370245	0.2625819
12	0.4497833	0.3926532	0.1575635
13	0.1598765	0.3946234	0.4455001
14	0.5337384	0.348435	0.1178265
15	0.6265731	0.2899149	0.0835119
16	0.5928456	0.3121286	0.0950258
17	0.7833934	0.1760466	0.04056
18	0.5323462	0.3492447	0.1184092
19	0.5169058	0.3580674	0.1250268
20	0.6024157	0.3059263	0.091658
21	0.5469929	0.3406158	0.1123913
22	0.5918246	0.3127853	0.0953901
23	0.7823876	0.1768215	0.0407909
24	0.5389312	0.3453955	0.1156733
25	0.4514354	0.3918849	0.1566797

注:p_{num}= 个人序号。
$p1$=“没有被剥夺”的概率。
$p2$=“轻度被剥夺”的概率。
$p3$=“严重被剥夺”的概率。

表 2.4 处于不同剥夺水平的次序 Probit 概率计算值:样本前 25 个观测值

P_{num}	$p1$	$p2$	$p3$
1	0.2564767	0.4197788	0.3237445
2	0.0561468	0.2607234	0.6831297
3	0.148693	0.3789889	0.4723181
4	0.2297609	0.4152423	0.3549968
5	0.2077654	0.4091158	0.3831188
6	0.0723101	0.2918726	0.6358173
7	0.3164209	0.4204631	0.263116
8	0.5960203	0.3162006	0.0877791
9	0.6006402	0.3134694	0.0858905
10	0.5609338	0.3361012	0.102965
11	0.3078035	0.4210712	0.2711253
12	0.4515077	0.3873188	0.1611736
13	0.1692549	0.3920921	0.438653
14	0.5286788	0.3530054	0.1183157
15	0.6201291	0.3016761	0.0781948
16	0.5868787	0.3215305	0.0915909
17	0.7813467	0.189157	0.0294963
18	0.5321902	0.3512332	0.1165766
19	0.5122991	0.3610432	0.1266577
20	0.5962939	0.3160395	0.0876666
21	0.5415704	0.3464162	0.1120134
22	0.5858997	0.3220953	0.092005
23	0.7803024	0.1899634	0.0297342
24	0.5342544	0.3501835	0.1155621
25	0.449369	0.3881331	0.1624979

注:p_{num}=个人序号。
$p1$=“没有被剥夺”的概率。
$p2$=“轻度被剥夺”的概率。
$p3$=“严重被剥夺”的概率。

对列出的 25 个人中的每个人来说,用 logit 和 probit 方法估计出的处于三种剥夺程度的概率并没有很大不同。从某种意义上说,用哪个模型并不重要,预测结果非常相似。

个人概率可用于生成剥夺程度的样本统计值。这些统

计值列于表 2. 5(logit 模型)和表 2. 6(probit 模型)中。

这两个表中显示的 $\hat{p}_{i1}$, $\hat{p}_{i2}$和 $\hat{p}_{i3}$的均值(分别用 $\bar{p}_1$, $\bar{p}_2$ 和 $\bar{p}_3$ 表示)在 logit 和 probit 模型间稍有不同。但是,取整数之后,这些不同消失了且估计值分别为:46%,35%和19%。在 logit 模型中,用上述方法算出的平均概率,等于在三个剥夺类别中的样本比例。[28] 这是次序 logit 模型的一个特性。在 probit 模型下平均概率很接近,但不等于样本比例,这种情况也是在实践中经常见到的。

表 2. 5 和表 2. 6 中的平均概率总和精确地为 1。这是因为个人的概率估计(平均概率从这些概率估计中计算而得)总和为 1。

$\hat{p}_{i1}$, $\hat{p}_{i2}$和 $\hat{p}_{i3}$的中位数(第 50 个百分位数),如表 2. 5 和表 2. 6 所示,在 logit 和 probit 模型间存在些微的差异,但是这些差异再次在取整数之后消失了,产生的中位数概率的估计值分别为 48%, 37%和 14%。对这两个模型中的任何一个来说,中位数概率都不一定总和为 1,在这个应用实例中,它们的总和确实并不等于 1。原因在于虽然对任何一个人估计值的总和为 1,但是 $\hat{p}_{i1}$, $\hat{p}_{i2}$和 $\hat{p}_{i3}$的中位数却不一定与同一个人相关。比如,如果 $\hat{p}_{i1}$, $\hat{p}_{i2}$和 $\hat{p}_{i3}$($i = 1, \cdots, 13164$)的值按照递增或递减的顺序排列,那么在 $\hat{p}_{i1}$排列中处于第 50 个百分位数的那个人可能与在 $\hat{p}_{i3}$中第 50 个百分位数的那个人不同。在 logit 的估计值方面(表 2. 5),没有被剥夺的概率($\hat{p}_{i1}$)处于最低的百分位数的是6. 7%,在这个等级里面最低的估计值为 1. 9%。在这个量表的另一端,$\hat{p}_{i1}$在最高的百分位数的值为 83. 1%,在这个级别里面最高的估计值为 87%。

表 2.5 不同剥夺程度的样本统计值:logit 模型

Pr(y = 1)				
	Percentiles	Smallest		
1%	0.0675202	0.0185615		
5%	0.1466481	0.0198162		
10%	0.1875675	0.0300967	Obs	13164
25%	0.303902	0.0347162	Sum of Wgt.	13164
50%	0.4843619		Mean	0.4592781
		Largest	Standard Deviation	0.1889648
75%	0.5993176	0.8697239		
90%	0.6954355	0.8697425	Variance	0.0357077
95%	0.7368268	0.8699151	Skewness	−0.1866478
99%	0.8308569	0.8699151	Kurtosis	2.155329
Pr(y = 2)				
	Percentiles	Smallest		
1%	0.1389147	0.0915195		
5%	0.2075179	0.0969696		
10%	0.239251	0.1077327	Obs	13164
25%	0.2989144	0.1077327	Sum of Wgt.	13164
50%	0.3674155		Mean	0.3493639
		Largest	Standard Deviation	0.076207
75%	0.4150781	0.4377985		
90%	0.4339717	0.4377985	Variance	0.0058075
95%	0.4366746	0.4377985	Skewness	−0.8337451
99%	0.4377566	0.4377985	Kurtosis	2.917122
Pr(y = 3)				
	Percentiles	Smallest		
1%	0.0301859	0.0223522		
5%	0.0517813	0.0223522		
10%	0.0627571	0.0223856	Obs	13164
25%	0.0927392	0.0223891	Sum of Wgt.	13164

续表

50%	0. 1399815		Mean	0. 191358
		Largest	Standard Deviation	0. 1405856
75%	0. 2593728	0. 8095671		
90%	0. 3984033	0. 8312855	Variance	0. 0197643
95%	0. 4708132	0. 8832142	Skewness	1. 451671
99%	0. 6786129	0. 889919	Kurtosis	4. 969932

表 2.6 不同剥夺程度的样本统计值:probit 模型

		Pr(y = 1)		
	Percentiles	Smallest		
1%	0. 0610129	0. 0089802		
5%	0. 1495682	0. 0095785		
10%	0. 1936998	0. 0190097	Obs	13164
25%	0. 311691	0. 0255515	Sum of Wgt.	13164
50%	0. 4870896		Mean	0. 4602124
		Largest	Standard Deviation	0. 1855043
75%	0. 5928672	0. 8747257		
90%	0. 6891779	0. 8747432	Variance	0. 0344118
95%	0. 7340055	0. 8749362	Skewness	−0. 212917
99%	0. 8299778	0. 8749362	Kurtosis	2. 250704
		Pr(y = 2)		
	Percentiles	Smallest		
1%	0. 1500103	0. 0957713		
5%	0. 2209043	0. 09959		
10%	0. 2547602	0. 1132009	Obs	13164
25%	0. 3088254	0. 1132009	Sum of Wgt.	13164
50%	0. 3662931		Mean	0. 3488479
		Largest	Standard Deviation	0. 0663654
75%	0. 4044376	0. 421618		
90%	0. 418923	0. 421618	Variance	0. 0044044
95%	0. 4207648	0. 421618	Skewness	−1. 039971
99%	0. 421588	0. 421618	Kurtosis	3. 528536

续表

Pr(y = 3)				
	Percentiles	Smallest		
1%	0.0194342	0.011863		
5%	0.0412402	0.011863		
10%	0.0542446	0.011892	Obs	13164
25%	0.0890826	0.0118946	Sum of Wgt.	13164
50%	0.1402645		Mean	0.1909397
		Largest	Standard Deviation	0.1432815
75%	0.2674853	0.7992974		
90%	0.4023936	0.8322631	Variance	0.0205296
95%	0.4708187	0.8908314	Skewness	1.315678
99%	0.668156	0.8952485	Kurtosis	4.483449

概率估计值:均值的估计

在前文中,平均概率的估计值从个人概率 $\hat{p}_{i1}$, $\hat{p}_{i2}$ 和 $\hat{p}_{i3}(i = 1, \cdots, 13164)$ 的估计值中计算得出。在 logit 模型中平均概率正好等于样本比例,在 probit 模型中,他们也很接近样本比例。但是,还有另外一种方法计算平均概率(见 Greene, 2000:879),且这种方法产生的结果与上面的不同。设 $\bar{X}_k = \sum_{i=1}^{N} X_{ik}/N$ 为第 k 个决定变量的平均值,让 $\hat{Z}_i$ 的均值等于:

$$\bar{Z} = \sum_{i=1}^{N} \hat{Z}_i/N = \sum_{i=1}^{N} \left(\sum_{k=1}^{K} \hat{\beta}_k X_{ik}\right)/N$$
$$= \sum_{k} \hat{\beta}_k \sum_{i} X_{ik}/N = \sum_{k} \hat{\beta}_k \bar{X}_k \qquad [2.17]$$

然后计算 $\hat{p}_1$, $\hat{p}_2$,和 $\hat{p}_3$ 如下:

$$\hat{p}_1 = \Pr(\varepsilon_i \leqslant \hat{\delta}_1 - \bar{Z}) \quad [2.18a]$$

$$\hat{p}_2 = \Pr(\hat{\delta}_1 - \bar{Z} < \varepsilon_i \leqslant \hat{\delta}_2 - \bar{Z}) \quad [2.18b]$$

$$\hat{p}_3 = \Pr(\varepsilon_i \geqslant \hat{\delta}_2 - \bar{Z}) \quad [2.18c]$$

通常，$\bar{p}_1 \neq \hat{p}_1$，$\bar{p}_2 \neq \hat{p}_2$，$\bar{p}_3 \neq \hat{p}_3$，这并不奇怪。因为尽管 $\bar{p}_j$ 和 $\hat{p}_j (j = 1, 2, 3)$两者都旨在测量处于不同剥夺水平的总体概率,但是它们的计算方式却非常不同。$\bar{p}_j$ 计算为估计到的个人概率 $\hat{p}_{ij}$ 的均值,而 $\hat{p}_{ij}$ 是从个人决定变量的值结合估计到的临界点计算出来的。另一方面，$\hat{p}_j$ 绕过了个人概率,反而是用个人 Z_i 的均值 $\bar{Z}$(或者,等于用决定变量 X_{ik} 的均值 $\bar{X}_k$ 计算而得),结合估计到的临界点而得出。表 2.7 和表 2.8 分别比较了 logit 和 probit 模型中 $\bar{p}_j$ 的值和 $\hat{p}_j$ 的值。

表 2.7 处于不同剥夺程度的总体概率:logit 模型

概　　率	计算为个人概率的均值:$\bar{p}$	从决定变量的均值计算而得:$\hat{p}$
没有被剥夺	0.459	0.447
轻度被剥夺	0.349	0.394
严重被剥夺	0.191	0.159

表 2.8 处于不同剥夺程度的总体概率:probit 模型

概　　率	计算为个人概率的均值:$\bar{p}$	从决定变量的均值计算而得:$\hat{p}$
没有被剥夺	0.460	0.452
轻度被剥夺	0.349	0.387
严重被剥夺	0.191	0.161

边际概率的计算:连续变量

模型中唯一的连续变量,如方程 2.16 所示,是 AGE_i 和

AGE_i^2。为了计算个人年龄每增加 1 年对他/她处于三种剥夺程度概率的影响，我们需要在相关点上估计合适的密度函数并乘以与两个变量相关的系数估计值(见方程 2.12a 至方程 2.12c，以及方程 2.13a 至方程 2.13c)。

还可以通过下列两个方式中的任何一种完成。

1. 对样本中的每个人，在相关点上对其计算边际效应。对个人 $i(i=1, \cdots, 13164)$，这些相关点为：$\hat{\delta}_1-\hat{Z}_i$ 对应于 $\Pr(Y_i=1)$ 的边际效应；$\hat{\delta}_2-\hat{Z}_i$ 和 $\hat{\delta}_1-\hat{Z}_i$ 对应于 $\Pr(Y_i=2)$ 的边际效应；$\hat{\delta}_2-\hat{Z}_i$ 对应于 $\Pr(Y_i=3)$ 的边际效应。用这些估计值乘以 AGE_i 的系数估计值($\hat{\beta}_4$)，同时也用这些估计值乘以 AGE_i^2 的系数估计值($\hat{\beta}_5$)。这两个结果相加就得到了每个人年龄的一个小变化(1 年)对他/她的概率的边际效应。然后所有这些个人作用的均值就产生了样本中每个人的年龄增加 1 年对(处于不同的剥夺程度的)概率的平均影响。

2. 计算样本中所有个人 $\hat{Z}_i$ 值的均值(见方程 2.17)。如果这个均值记为 $\bar{Z}$，在相关点计算边际效应：$\hat{\delta}_1-\bar{Z}$ 对应于 $\Pr(Y_i=1)$ 的边际效应；$\hat{\delta}_2-\bar{Z}$ 和 $\hat{\delta}_1-\bar{Z}$ 对应于 $\Pr(Y_i=2)$ 的边际效应；$\hat{\delta}_2-\bar{Z}$ 对应于 $\Pr(Y_i=3)$ 的边际效应。

用这些估计值乘以 AGE_i 的系数估计值，同时也用这些估计值乘以 AGE_i^2 的系数估计值。这两个结果相加就得到了每个人年龄的一个小变化(1 年)对他/她的概率的边际效应。然后所有这些个人作用的均值就产生了样本中所有人的平均年龄增加 1 年对(处于不同的剥夺程度的)概率的影响。

表 2.9 和表 2.10 展示了用以上两种不同的方法计算出来的 logit 和 probit 模型中年龄增加的边际效应。

表 2.9　年龄对处于不同剥夺程度的边际效应:logit 模型

概　率	计算为个人边际效应的均值	从决定变量的均值计算而得
没有被剥夺	−0.00523	−0.00607
轻度被剥夺	0.00191	0.00279
严重被剥夺	0.00332	0.00329

表 2.10　年龄对处于不同剥夺程度的边际效应:probit 模型

概　率	计算为个人边际效应的均值	从决定变量的均值计算而得
没有被剥夺	−0.00524	−0.00591
轻度被剥夺	0.00171	0.00226
严重被剥夺	0.00353	0.00365

边际概率的计算:虚拟变量

在前面我们观察到一个虚拟变量的作用应该通过比较它取其中一个值时的概率与取另一个值的概率进行分析,在这个比较过程中其他变量的值保持不变。这个方法现在用于分析宗教信仰对处于不同剥夺程度的概率的影响,通过比较(对所有的 i)$CT_i = 1$(宗教虚拟变量)的情况与(对所有的 i)$CT_i = 0$ 的情况。与连续变量的情形一样,这种广义的方法可以采用以下两种途径中的任意一种。

1. 假设其他条件均同且样本中的所有人都是天主教徒 $CT_i = 1(i = 1\cdots 13164)$。那么在这种情况下,$\hat{Z}_i$($Z_i$ 的估计值)从方程 2.16 中用系数估计值以及对所有 $iCT_i = 1$ 计算而得。把这个估计出来的值叫做 $\hat{Z}_i^c$。

让 $\hat{p}_{ij}^{c}$ 表示在这个假设的情况下个人 i 处于剥夺程度 j ($j = 1, 2, 3$)的(计算出来的)概率,其中这些概率从方程 2.5a至 2.5c 中——或者,如果使用 probit 模型,从方程 2.11a 至方程 2.11c 中——用 $\hat{Z}_i = \hat{Z}_i^c$ 计算而得。现在假设其他条件相同且样本中的所有人都是新教徒,对所有 $i = 1 \cdots 13164$, $CT_i = 0$,让 $\hat{Z}_i^p$ 表示 Z_i 的估计值。从方程 2.6 可知,$\hat{Z}_i^c = \hat{Z}_i^p + \hat{\beta}_3$。让 $\hat{p}_{ij}^{p}$表示在这些假设的情况下个人 i 处于剥夺程度 j ($j = 1, 2, 3$)的(计算出来的)概率。对任何样本中的个人 i,$\hat{p}_{ij}^{c}$和 $\hat{p}_{ij}^{p}$的差别完全来自于宗教信仰的差别:$\hat{p}_{ij}^{c}$ 是用 $\hat{Z}_i^c$——也就是,天主教徒这个变量等于 1 时——计算出来的;$\hat{p}_{ij}^{p}$是用 $\hat{Z}_i^p$,天主教徒这个变量等于 0 时计算出来的;在其他变量的值没有任何变化的情况下。如果 $\bar{p}_j^c = \sum_{i=1}^{N} \hat{p}_{ij}^{c}/N$ 和 $\bar{p}_j^p = \sum_{i=1}^{N} \hat{p}_{ij}^{p}/N$ 分别是两组概率估计值的均值,那么它们之间的差异就测量出了宗教信仰对处于不同剥夺程度的平均概率的影响。

2. 另一个方法是比较对样本中的所有人,变量 CT_i 取两个不同值时的两种概率,且在每种情况下其他变量都取样本均值。让 $\bar{Z}^c$ 和 $\bar{Z}^p$ 分别表示对所有 $i = 1 \cdots 13164$, $CT_i = 1$ 和 $CT_i = 0$ 时 $\bar{Z}$ 的值,且在每一种情况下其他变量 $X_k = \bar{X}_k$。由方程 2.16,$\bar{Z}^c = \bar{Z}^p + \hat{\beta}_3$。实际上,这个应用除了宗教信仰外还包括了样本的平均特征,而且在一种情况下它是天主教徒,在另一种情况下它是新教徒。那么,在方程 2.5a 至方程 2.5c 中使用 $\bar{Z}^c$,然后 $\bar{Z}^p$,就可以对这两种假定的情形(前一种情形下这个人为天主教徒,后者这个人为新教徒)计算出(处于没有被剥夺、轻度被剥夺、严重被剥夺的)三种概

率的估计值。如果这两种概率分别记为 $\hat{q}_j^c$ 和 $\hat{q}_j^p(j=1, 2, 3)$,那么它们之间的差异就测量出了宗教信仰对处于不同剥夺程度的概率的影响。

表2.11展示了在logit模型下 $\bar{p}_j^c$ 和 $\bar{p}_j^p$,以及 $\hat{q}_j^c$ 和 $\hat{q}_j^p$ $(j=1, 2, 3)$的值,表2.12则表示probit模型下的情况。这两个表有两个重要的特点。首先,对于两种计算边际效应的方法中的任何一种,logit和probit概率之间都几乎没有差异。其次,对任何一个模型来说,用两种不同的方法计算出来的概率之间存在着非常大的差异。注意,用个人概率的均值计算出来的处于严重被剥夺的概率显著地高于用样本的平均特征计算出来的处于严重被剥夺的概率。

表2.11 宗教信仰对处于不同剥夺程度的概率的影响:logit模型

概　率	计算为个人边际效应的均值		从决定变量的均值计算而得	
	CAT	*PRT*	*CAT*	*PRT*
没有被剥夺	0.432	0.472	0.417	0.462
轻度被剥夺	0.360	0.346	0.407	0.387
严重被剥夺	0.207	0.182	0.176	0.151

表2.12 宗教信仰对处于不同剥夺程度的概率的影响:probit模型

概　率	计算为个人边际效应的均值		从决定变量的均值计算而得	
	CAT	*PRT*	*CAT*	*PRT*
没有被剥夺	0.433	0.473	0.421	0.466
轻度被剥夺	0.359	0.346	0.398	0.382
严重被剥夺	0.208	0.191	0.181	0.152

处于严重被剥夺的状态是拥有剥夺决定变量的“极端”值的结果。极端值的影响在个人其他变量值被设置为样本均值时被减弱了。从另一方面来说,在概率计算中使用个人

的变量值时，极端值被允许充分地发挥了作用。

那么哪一种方法更适合使用呢？关键的问题是，当我们感兴趣的虚拟变量取各不相同的两种值的时候，其他变量的值是如何保持不变的。实际上，第二种方法对每个人都分配了样本的平均值。但是，平均值并不是神圣不可侵犯的，这个分配的共同值也可以用中位数或者任意其他值，且这些不同分配方式中的每一种都会在计算出来的概率 $\hat{q}_j$ 上产生一个不同的结果。

第一种方法就没有这个缺陷。因为个人的值没有被干扰，计算出来的个人的概率是唯一的，所以在两种不同的情形下，平均[29]概率 $\bar{p}_j$ 是唯一的。

第4节 | 对次样本的估计:特征与系数

宗教信仰由于方程2.16中β_3的存在而对处于不同剥夺程度的概率(表2.11和表2.12)有影响。这意味着,在第一种方法下$\hat{Z}_i^c \neq \hat{Z}_i^p$,在第二种方法下$\bar{Z}^c \neq \bar{Z}^p$。因此,样本中的每个人$CT_i = 1$时被剥夺的概率不同于样本中每个人$CT_i = 0$时的概率。但这导致我们考虑这样一种可能性,即对方程2.16中的每个变量,"天主教徒"的系数都不同于"新教徒"的系数。也就是说,方程2.16的估计应该允许天主教徒与新教徒的系数互不相同。有两种(几乎等同的)的操作方法。第一种只估计一个方程,但允许方程中的每个天主教徒的系数都与相应的新教徒的系数不同。第二种对天主教徒和新教徒次样本分别估计两个不同的方程。为了执行第一种策略,定义方程如下:

$$
\begin{aligned}
D_i = & \beta_1 + \gamma_1 * CT_i + \beta_2 * SEX_i + \gamma_2 * CT_i * SEX_i + \beta_3 * AGE_i \\
& + \gamma_3 * CT_i * AGE_i + \beta_4 * AGE_i^2 + \gamma_4 * CT_i * AGE_i^2 \\
& + \beta_5 * HIGHED_i + \gamma_5 * CT_i * HIGHED_i + \beta_6 * MIDED_i \\
& + \gamma_6 * CT_i * MIDED_i + \beta_7 * RET_i + \gamma_7 * CT_i * RET_i \\
& + \beta_8 * INAC_i + \gamma_8 * CT_i * INAC_i + \beta_9 * UE_i
\end{aligned}
$$

$$+\gamma_9 * CT_i * UE_i + \beta_{10} * HNUM_i + \gamma_{10} * CT_i * HNUM_i$$
$$+\beta_{11} * SNPAR_i + \gamma_{11} * CT_i * SNPAR_i + \beta_{13} * AREA_{2i}$$
$$+\cdots+\beta_{21} * AREA_{10i} + \gamma_{13} * CT_i * AREA_{2i} + \cdots$$
$$+\gamma_{21} * CT_i * AREA_{10i} + \varepsilon_i = Z_i + Z_i' + \varepsilon_i \qquad [2.19]$$

通过在方程中加入“交互变量”——即用方程 2.16 中的原解释变量乘以天主教徒虚拟变量——某个特定变量对剥夺的影响就可依据受关注的人是新教徒还是天主教徒而不同。在方程 2.19 中，β_k 是“新教徒”的系数[30]，而 γ_k——交互变量的系数——代表了作为天主教徒而导致的对这些系数的附加贡献。因此 γ_k 的值测量了“交互作用”的强度。所以对 $\gamma_k=0$ 的检验就是对第 k 个解释变量与个人的宗教信仰之间没有交互影响这个零假设的检验。换一种说法，$\gamma_k=0$ 意味着在第 k 个解释变量上新教徒与天主教徒的系数之间没有差别。与两个不同方程的方法相比，这个单一方程“整合”方法的一大优势在于，它可以轻易地检验某个变量对这两组人是否有不同的影响。实际上，随后的内容还对这个单一方程“整合”方法进行了一定程度的探讨。但是，为了突出说明第二种方法，方程 2.16 对天主教徒和新教徒次样本分别进行了估计。表2.13和表 2.14 分别展示了当误差被假设为逻辑分布时，方程 2.16 对天主教徒和新教徒的估计结果。[31]

在表 2.13 和表 2.14 所示的方程估计值中，变量 *dwn* 和 *dry* 被丢掉了。这是因为它们的系数单个地（在 Z-score 上）以及共同地（在似然比检验上）都不显著地不等于 0。这是一个方法上的重要问题：当方程用于预测时，它们应该包含所有的变量，即使其中的一些系数可能并不显著地不等

于 0,还是应该只包括那些具有显著地不等于 0 的系数的变量。

表 2.13 北爱尔兰社会剥夺程度的次序 logit 模型结果:新教徒样本

Ordered Logit Estimates Log Likelihood = −8124.0888; *Number of obs* = 8800;
LR $\chi^2(17)$ = 1527.75; *Prob*>χ^2 = 0.0000; *Pseudo* R^2 = 0.0859

y	*Coefficient*	*Standard Error*	*z*	*P*>|*z*|	[95% *Conf.*	*Interval*]
sex	−0.2293014	0.0444134	−5.163	0.000	−0.31635	−0.1422527
age	−0.0275073	0.0060613	−4.538	0.000	−0.0393873	−0.0156273
age2	0.0008768	0.0001131	7.754	0.000	0.0006551	0.0010984
ret	0.6275465	0.1008167	6.225	0.000	0.4299493	0.8251437
inac	1.179892	0.0954488	12.362	0.000	0.9928158	1.366968
ue	1.556987	0.0848854	18.342	0.000	1.390615	1.723359
highed	−0.8085732	0.085388	−9.469	0.000	−0.9759306	−0.6412159
mided	−0.6320822	0.1490549	−4.241	0.000	−0.9242244	−0.33994
hnum	0.95203	0.08401	11.332	0.000	0.7873735	1.116686
snpar	0.3647026	0.0845436	4.314	0.000	0.1990001	0.5304051
ard	−0.4509469	0.063093	−7.147	0.000	−0.5746069	−0.327287
crk	−0.3349415	0.0715481	−4.681	0.000	−0.4751733	−0.1947098
ant	0.3530984	0.0758826	4.653	0.000	0.2043712	0.5018256
col	0.368263	0.0833618	4.418	0.000	0.2048769	0.5316491
arm	0.4664409	0.0988147	4.720	0.000	0.2727676	0.6601141
ban	0.2097729	0.076176	2.754	0.006	0.0604708	0.359075
frm	0.6888737	0.0980013	7.029	0.000	0.4967947	0.8809528
_cut1	0.1488535	0.0834433		(Ancillary parameters)		
_cut2	2.04582	0.0869085				

y	*Probability*	*Observed*
1	Pr($xb+u<$ _cut1)	0.4970
2	Pr(_cut1 $<xb+u<$ _cut2)	0.3399
3	Pr(_cut2 $<xb+u$)	0.1631

表 2.14 北爱尔兰社会剥夺程度的次序 logit 模型结果：天主教徒样本

Ordered Logit Estimates Log Likelihood = −4283.363; *Number of obs* = 4364; $LR\chi^2(17)$ = 875.06; $Prob>\chi^2$ = 0.0000; *Pseudo* R^2 = 0.0927						
y	*Coefficient*	*Standard Error*	*z*	$P>\|z\|$	[95% *Conf.*	*Interval*]
sex	−0.0430983	0.0621282	−0.694	0.488	−0.1648672	0.0786707
age	−0.0181642	0.0089546	−2.028	0.043	−0.0357148	−0.0006136
age2	0.00064	0.0001754	3.648	0.000	0.0002961	0.0009839
ret	0.6109521	0.1773859	3.444	0.001	0.2632821	0.958622
inac	1.487517	0.1299683	11.445	0.000	1.232784	1.742251
ue	1.364899	0.0906986	15.049	0.000	1.187133	1.542665
highed	−1.09291	0.1133639	−9.641	0.000	−1.315099	−0.8707208
mided	−0.7667852	0.2172056	−3.530	0.000	−1.1925	−0.3410701
hnum	0.9087931	0.0739094	12.296	0.000	0.7639333	1.053653
snpar	0.239649	0.0987062	2.428	0.015	0.0461884	0.4331096
ard	−0.5853662	0.16575	−3.532	0.000	−0.9102303	−0.2605022
crk	−0.3226664	0.1620424	−1.991	0.046	−0.6402637	−0.0050692
ant	0.2364382	0.1329838	1.778	0.075	−0.0242052	0.4970816
col	0.4795002	0.1039698	4.612	0.000	0.2757232	0.6832773
arm	0.2546794	0.0926459	2.749	0.006	0.0730968	0.4362621
ban	0.2025872	0.0954257	2.123	0.034	0.0155563	0.389618
frm	0.5243653	0.0950858	5.515	0.000	0.3380006	0.7107301
_cut1	0.0110645	0.1138485		(Ancillary parameters)		
_cut2	1.866844	0.1178967				

y	*Probability*	*Observed*
1	Pr($xb+u<$ _cut1)	0.3824
2	Pr(_cut1 $<xb+u<$ _cut2)	0.3682
3	Pr(_cut2 $<xb+u$)	0.2493

一种观点是，如果我们相信某个变量在方程设定中有一个合法的位置，那我们就应该坚持这个想法且不管怎样都把这个变量包括在内。但另一种观点是，因为估计与预测的目

的是为了使方程设定面对数据,把预测基于从完整的模型设定中得到的系数估计值有可能具有误导性,因为它可能允许那些在模型设定中其合法性被数据明显"拒绝"的变量影响我们的预测。在前文中,追随第一种观点,预测基于完整的模型设定;在这一部分中,追随第二种观点,预测基于受限制的模型设定。

表2.15说明:新教和天主教样本分别有50%和38%的人没有被剥夺,34%和37%的人轻度被剥夺,16%和25%的人严重被剥夺。与新教徒相比,天主教徒没有被剥夺的比例更小,轻度和严重被剥夺的比例更高(见表2.15),这种实际情况可能有两个方面的原因:第一,那些使被剥夺的概率增加的特征不均衡地集中于天主教徒,以及/或者那些使被剥夺的概率下降的特征不均衡地集中于新教徒;[32]第二,某个特定的属性对天主教徒来说,被更加严厉地惩罚(如果是能使剥夺增加的:比如处于失业状况),以及/或者得到的回报更少(如果是能使剥夺降低的:比如具有教育资格)。

表2.15 北爱尔兰社会剥夺程度:按宗教类别划分

	样本中的百分比		
	没有被剥夺	轻度被剥夺	严重被剥夺
总　体	45.9	34.9	19.2
天主教徒	38.2	36.8	24.9
新教徒	49.7	34.0	16.3

为了确定天主教—新教之间的剥夺差距(定义为在不同的剥夺程度上天主教徒与新教徒的比例差异)在多大程度上是由特征差异造成的,在多大程度上是由系数差异造成的,在下列问题上提出了计量经济学方面的议题。

1. 如果新教徒与天主教徒用各自的系数估计值评估他们具有的特征，那么对他们处于不同剥夺程度的预测概率会是怎样的？把这些概率分别记为 $\hat{p}_{ij}^{p}$ 和 $\hat{p}_{ij}^{c}$（$i=1, \cdots, 13164$；$j=1, 2, 3$），它们的均值记为 $\bar{p}_{j}^{p}$ 和 $\bar{p}_{j}^{c}$（$j=1, 2, 3$）。[33]

2. 如果新教徒的特征用天主教徒的系数估计值进行评估，那么它们处于不同的剥夺程度的预测概率会是怎样的呢？把这些“合成的”概率记为 $\hat{q}_{ij}^{p}$（$i=1, \cdots, 13164$；$j=1, 2, 3$），它们的均值记为 $\bar{q}_{j}^{p}$（$j=1, 2, 3$）。

3. $\bar{q}_{j}^{p}$ 与 $\bar{p}_{j}^{c}$ 合 $\bar{p}_{j}^{p}$ 相比怎样呢？

表2.16展示了作为一个整体时，以及在各自的次样本中，这两组人的 $\bar{p}_{j}^{p}$，$\bar{q}_{j}^{p}$ 与 $\bar{p}_{j}^{c}$ 的值这三个概率。表2.16说明，对样本中所有个人进行计算，新教徒处于没有被剥夺，轻度被剥夺和严重被剥夺的平均“自己系数的”[34]概率分别为50%，34%和16%。当新教徒的特征用天主教徒的系数进行评估时，没有被剥夺的概率下降到46%，严重被剥夺和轻度被剥夺的概率分别上升到19%和35%。这个故事在其他被分析的次样本中是类似的（单身父/母亲、退休者、不从事经济活动者、失业者以及住在大家庭中的人）。当新教徒的特征用天主教徒的系数进行评估时，除了两个例外（退休者和失业者），没有被剥夺的概率总是下降，严重被剥夺的概率总是上升，虽然很自然地，这些变化的大小随着考虑的次样本的不同而变化。[35]（在没有被剥夺的概率中的）最大的下降与（在强烈被剥夺的概率中的）最明显的上升出现在那些不从事经济活动的人当中。

表 2.16 天主教徒和新教徒处于不同剥夺程度的概率估计值*

	概率估计值		
	没有被剥夺	轻度被剥夺	严重被剥夺
总体			
新教徒(用新教徒的系数估计)	49.7	34.0	16.2
新教徒(用天主教徒的系数估计)	45.8	35.3	18.9
天主教徒(用天主教徒的系数估计)	38.2	36.9	24.9
单身父/母亲			
新教徒(用新教徒的系数估计)	37.2	38.2	24.6
新教徒(用天主教徒的系数估计)	35.0	38.1	26.9
天主教徒(用天主教徒的系数估计)	27.3	37.8	34.8
退休者			
新教徒(用新教徒的系数估计)	24.8	41.4	33.8
新教徒(用天主教徒的系数估计)	24.6	40.7	34.6
天主教徒(用天主教徒的系数估计)	20.7	39.6	39.7
不从事经济活动者			
新教徒(用新教徒的系数估计)	31.3	41.4	27.3
新教徒(用天主教徒的系数估计)	20.7	38.9	40.5
天主教徒(用天主教徒的系数估计)	16.5	37.6	45.9
失业者			
新教徒(用新教徒的系数估计)	21.1	40.3	38.6
新教徒(用天主教徒的系数估计)	21.3	39.5	39.2
天主教徒(用天主教徒的系数估计)	16.3	36.3	47.4
居住于大家庭者			
新教徒(用新教徒的系数估计)	28.1	40.0	31.9
新教徒(用天主教徒的系数估计)	26.0	39.0	34.9
天主教徒(用天主教徒的系数估计)	23.2	38.6	38.2

注：* 用表 2.13 和表 2.14 的次序 logit 估计结果计算而得。

在轻度被剥夺的概率方面，如果新教徒的特征用天主教徒的系数进行评估，在单身父/母亲，或退休者，或失业者，或居住在大家庭的人中，新教徒的位置不变，[36]但在那些不从事经济活动者中情况变差了。

新教徒和天主教徒在不同的剥夺类别中比例的差别，$\hat{p}_j^p - \hat{p}_j^c$ 可以分解为：

$$\bar{p}_j^p - \bar{p}_j^c = (\bar{q}_j^p - \bar{p}_j^c) + (\bar{p}_j^p - \bar{q}_j^p) = A_j^* + B_j^* \quad [2.20]$$

在方程 2.20 中，A_j^* 表示在天主教徒与新教徒的“剥夺差异”中因这两组人之间特征的差别而导致的部分，B_j^* 表示由系数值的组间差别而导致的部分。这些绝对差异可以用比例表示：

$$A_j = A_j^* / (\bar{p}_j^p - \bar{p}_j^c) \text{ 以及 } B_j = 1 - A_j \quad [2.21]$$

表 2.17 用百分比的形式，展示了 A_j 和 B_j 的值。

表 2.17 说明，对样本中所有个人进行计算，在那些没有被剥夺的天主教徒与新教徒之间，66%的剥夺差异（$\bar{p}_j^p - \bar{p}_j^c$）是由于天主教次样本中的人具有与新教次样本中的人不相同的特征而引起的，34%是因为天主教徒特征的评估方式不同于新教徒特征的评估方式。至于严重被剥夺的情况，天主教徒—新教徒之间的差异有 69%是由特征的不同造成的，但是在轻度被剥夺方面，只有 55%的天主教徒—新教徒差异来源于特征差别。

转向具体的分组，对那些单身父/母亲、退休者、失业者来说，在不同的剥夺程度上，相对来说较大比例的天主教徒—新教徒差异来自于两组人之间的特征差别。相对来说较小比例的差异是因为具体的一些特征，在应用到天主教徒身上时比应用到新教徒时具有更加严重的剥夺后果。

表 2.17　对天主教徒和新教徒间剥夺程度差异的贡献率*

	百分比贡献率	
	特征差异(A_j)	系数差异(B_j)
总体		
没有被剥夺	65.8	34.2
轻度被剥夺	54.7	45.3
严重被剥夺	69.4	30.6
单身父/母亲		
没有被剥夺	78.2	21.8
轻度被剥夺	64.4	35.6
严重被剥夺	77.7	22.3
退休者		
没有被剥夺	96.1	3.9
轻度被剥夺	63.1	36.9
严重被剥夺	85.7	14.3
不从事经济活动者		
没有被剥夺	28.3	71.7
轻度被剥夺	33.9	66.1
严重被剥夺	29.4	70.6
失业者		
没有被剥夺	103.0	−3.0
轻度被剥夺	81.7	18.3
严重被剥夺	93.5	6.5
居住于大家庭者		
没有被剥夺	58.1	41.9
轻度被剥夺	27.2	72.8
严重被剥夺	51.4	48.6

注：* 用表 2.16 的概率估计值计算而得。

比如，对失业者来说，在严重被剥夺方面，94%的天主教徒—新教徒差异可以被这两组人之间特征的不同来解释。

但是，对不从事经济活动者来说，被用于评估特征的系数的差别（相对于特征本身的差异）解释了超过 2/3 的天主教徒与新教徒之间的剥夺差异。对于住在大家庭中的人来说，在严重被剥夺方面几乎一半的组间差异，以及在轻度被剥夺方面超过 70％的组间差异，是由用于评估剥夺来源特征的系数的不同而引起的。

第3章

多类别模型

第1节 | 简介

前面考查了一组围绕着多元（>2）结果事件的模型，且这些结果在本质上是次序的。在这一系列模型里面，因变量Y_i，在对个人i界定这些结果时（对第一种结果$Y_i=1$；对第二种结果$Y_i=2$；以此类推，直到对第M个（最后一个）结果$Y_i=M$；$i=1, \cdots, N$），是一个离散的、有序的变量。估计这些模型的合适的方法是次序 logit 和次序 probit。这一部分关注的是结果没有次序的多元结果模型。多类别 logit 的方法论——对这组模型的适合的估计方法——在对职业选择的分析中进行解释并加以应用。多类别 logit 的一个重要特征（和局限）是无关选项的独立性（IIA）。这一部分的结尾对这个特征以及怎样可能回避它的局限进行了讨论。

一些真实事件为我们提供了非次序结果的例子。交通工具（乘坐巴士、火车或者小汽车）的选择就是一个明显的例子。其他非次序结果的例子包括职业选择（非技术的、技术的、专业的、管理的），居住地的选择（北边、南边、东边、西边），以及选举时的政党选择（保守党、自由党、工党）。上述例子的一个共有的词语就是"选择"。因为这些结果并没有什么本质上的好或者坏，因此可以认为个人是从可得到的结果的选项中选择最合适他们的结果。其实，效用最大化的框架为理解非次序多元结果模型的结构提供了一个很好的切入点。

第2节 | 随机效用模型

给定M个选项(标记为$j = 1, \cdots, M$)中的一个选择,第i个人从第j个选项中得出的效用可以表示为U_{ij}。假设这个效用是H个因素(决定变量)的一个线性函数。在这H个因素中,假设有R个是针对个人且与选择的属性无关的,S个($H = R + S$)是针对选择而与个人无关的。比如,在选择上班的交通方式方面,一个个人特征可能是他/她并不住在火车站附近。这个人几乎不坐火车去上班的事实与火车这个交通方式的属性没有关系,而仅仅是这个人住在哪里的结果。另一方面,高峰期交通非常繁忙这个实际情况是汽车出行的一个属性,它降低了所有人坐汽车上班的效用。假设代表了第i个人的特征的这R个变量的值为X_{ir}, $r = 1, \cdots, R$,代表了第j种选择的属性的变量的值为W_{js}, $s = 1, \cdots, S$。效用函数可以写成:

$$U_{ij} = \sum_{r=1}^{R} \beta_{jr} X_{ir} + \sum_{s=1}^{S} \gamma_{is} W_{js} + \varepsilon_{ij} = Z_{ij} + \varepsilon_{ij} \qquad [3.1]$$

其中β_{jr}是对第j种选择与第r个特征($r = 1, \cdots, R$)相关的系数,γ_{is}是对第i个人与第s种属性($s = 1, \cdots, S$)相关的系数。而且:

$$Z_{ij} = \sum_{r=1}^{R} \beta_{jr} X_{ir} + \sum_{s=1}^{S} \gamma_{is} W_{js} \qquad [3.2]$$

如果 $\beta_{jr} > 0$，个人 i 第 r 个特征的值 X_{ir} 的增加，会导致他/她从选择 j 中得到的效用上升，如果 $\beta_{jr} < 0$ 则下降。如果 $\gamma_{is} > 0$，选择 j 的第 s 种属性的值 W_{js} 的增加，对个人 i 会导致效用的上升，如果 $\gamma_{is} < 0$ 则下降。但是，因为效用与它的决定变量之间的关系并不精确——比如，可能有些因素被排除在方程之外或者有些因素的测量不准确——方程中加入了一个误差项 ε_i 来捕捉这种不精确性。所以称为随机效用模型。

当而且只有当在所有可利用的选择中，第 m 种选择提供了最高程度的效用时，一个人才会选择 $j = m$。也就是说，如果 Y_i 是一个随机变量，它的值 $j = 1, \cdots, M$ 表示个人 i 所做的选择，那么个人 i 会选择选项 m 的概率为：

$$\begin{aligned}
\Pr(Y_i = m) &= \Pr(U_{im} > U_{ij}) \quad \text{对所有 } j = 1, \cdots, M, \text{且 } j \neq m \\
&\Rightarrow \Pr(Z_{im} + \varepsilon_{im} > Z_{ij} + \varepsilon_{ij}) \\
&\Rightarrow \Pr(\varepsilon_{ij} - \varepsilon_{im} < Z_{im} - Z_{ij}) \quad \text{对所有 } j = 1, \cdots, M, \text{且 } j \neq m
\end{aligned}$$

第3节 | logit 模型的类别：多类别 logit 与条件 logit

麦克法登(McFadden, 1973)已经证明过，如果 M 个误差项 ε_{ij} ($j = 1, \cdots, M$)是以威布尔(Weibull)分布 $F(\varepsilon_{ij}) = \exp[\exp(-\varepsilon_{ij})]$ 独立同分布的，那么，

$$\Pr(Y_i = m) = \frac{\exp(Z_{im})}{\sum_{j=1}^{M}\exp(Z_{ij})} \qquad [3.3]$$

用方程 3.3 定义不同结果 $j = 1, \cdots, M$ 的概率的模型在这里称为广义 logit 模型。“广义”这个词表示模型同时包含了特征的作用和属性的作用，在方程 3.1 中分别为 X_{ir} 和 W_{js}。在广义 logit 模型这个类别里面，还可以再细分为两小类：

1. 多类别 logit 模型。这些模型只包含特征作用，所以对于它们来说方程 3.1 中所有的 $\gamma_{is} = 0$. 实际上，这些模型在个人层次的(individual specific)数据中使用。

2. 条件 logit 模型。这些模型只包含属性作用，所以对于它们来说方程 3.1 中所有的 $\beta_{jr} = 0$。实际上，这些模型在针对选项的(choice specific)数据中使用。

第 4 节 | 多类别 logit 模型

这个多类别模型由方程 3.3 定义但附带条件。现在，由于 $\gamma_{is}=0$，

$$Z_{ij}=\sum_{r=1}^{R}\beta_{jr}X_{ir} \qquad [3.4]$$

因为 $\Pr(Y_i=j)$ 概率对所有的选择来说总和为 1（即 $\sum_{j=1}^{M}\Pr(Y_i=j)=1$），只有 $M-1$ 个概率可以独立决定。因此，方程 3.3 中的多类别 logit 模型是不定式，因为它是一个具有 M 个方程但只有 $M-1$ 个独立未知数的体系。解决这个问题的实用的标准化方法是设定 $\beta_{1r}=0$，$r=1, \cdots, R$。

在这个标准化下，$Z_{i1}=0$，所以由方程 3.3：

$$\Pr(Y_i=1)=\frac{1}{1+\sum_{j=2}^{M}\exp(Z_{ij})} \qquad [3.5a]$$

$$\Pr(Y_i=m)=\frac{\exp(Z_{im})}{1+\sum_{j=2}^{M}\exp(Z_{ij})} \quad m=2, \cdots, M \qquad [3.5b]$$

这个标准化的结果是，这些概率是唯一决定的，所以通过所采用的标准化方法，在方程 3.5a 已经确定了 $\Pr(Y_i=1)$后，方程 3.5b 表示一个具有 $M-1$ 个方程和 $M-1$ 个独立未知数的概率 $\Pr(Y_i=m)$的体系。

从方程 3.5a 和方程 3.5b 中，结果 $j = k$ 与结果 $j = m$ 的概率比的对数为：

$$\log\left(\frac{\Pr(Y_i = m)}{\Pr(Y_i = k)}\right) = \sum_{r=1}^{R}(\beta_{mr} - \beta_{kr})X_{ir} = Z_{im} - Z_{ik} \quad [3.6]$$

那么这个风险比(risk-ratio)的对数(即结果 m 与结果 k 的概率比的对数，或者 $\log[\mathrm{Prob}(Y_i = m)/\mathrm{Prob}(Y_i = k)]$)并不取决于其他的选择。这个“风险比”，或者有时候称为相对风险—— $\mathrm{Prob}(Y_i = m)/\mathrm{Prob}(Y_i = k)$ ——可以很容易地通过对风险比的对数取指数而计算出来。如果 $k = 1$，那么风险比的对数为：

$$\log\left(\frac{\Pr(Y_i = m)}{\Pr(Y_i = 1)}\right) = \sum_{r=1}^{R}\beta_{mr}X_{ir} = Z_{im} \quad (m = 2, \cdots, M)$$

[3.7]

而风险比则为：

$$\frac{\Pr(Y_i = m)}{\Pr(Y_i = 1)} = \exp\left(\sum_{r=1}^{R}\beta_{mr}X_{ir}\right) = \exp(Z_{im}) \quad (m = 2, \cdots, M)$$

[3.8]

要把这个风险比(RR)与比数比(OR)区别开来，后者指的是一个结果的概率除以 1 减去那个结果的概率。

因此，对 $j = m$，比率比为：

$$OR_m = \frac{\Pr(Y_i = m)}{1 - \Pr(Y_i = m)} = \frac{\Pr(Y_i = m)}{\Pr(Y_i = 1)} \frac{\Pr(Y_i = 1)}{1 - \Pr(Y_i = m)}$$

$$= \frac{RR_m \Pr(Y_i = 1)}{1 - RR_m \Pr(Y_i = 1)} \quad [3.9]$$

其中，OR_m 和 RR_m 分别为与结果 $j = m$ 相关的比率比和风险比(后者以“基准”结果 $j = 1$ 为参照)。

在二分模型中，*RR* 与 *OR* 没有区别，因为基准结果 $Y_i = 1$ 就是 $Y_i \neq m$。在大于两种可能结果的模型中，结果 $Y_i = 1$ 与 $Y_i \neq m$ 是不一样的。所有 logit 模型的固有方法就是用一个结果与某个基准结果的可能性的比率表示分析结果——也就是，计算 *RR*。但是在二分模型中，*RR* 就是 *OR*，所以在这种模型中分析结果是用后者表示的。另一方面，在多类别 logit 模型中，由于 *RR* 与 *OR* 不一样，因此分析结果用 *RR* 表示。

估计与预测

因变量 $Y_i(i = 1, \cdots, N)N$ 种观测值中的每一个都视为从一个有 M 个结果的多项分布中的一次单一的抽取。定义虚拟变量 $\delta_{ij} = 1$，如果个人 i 选择 j，否则 $\delta_{ij} = 0$，$j = 1, \cdots, M$。则观察到这个样本的概率为：

$$L = \prod_{i=1}^{N}\prod_{j=1}^{M}[\Pr(Y_i = j)]^{\delta_{ij}} \Rightarrow \log L = \sum_{i=1}^{N}\sum_{j=1}^{M}\delta_{ij}\Pr(Y_i = j) \qquad [3.10]$$

其中 $\Pr(Y_i = j)$ 如果 $j = 1$ 的话由方程 3.5a 定义，如果 $j > 1$ 的话由方程 3.5b 定义。参数估计值 $\hat{\beta}_{jr}(j = 1, \cdots, M, r = 1, \cdots, R)$ 的选择是为了使似然函数最大化(方程 3.10)。

给定这些估计值，对每个个人 $i = 1, \cdots, N$，我们可以使用方程 3.4，对每种结果 $j = 1, \cdots, M$ 用 $\hat{\beta}_{jr}$ 代替 β_{jr}，产生 Z_{ij} 的估计值。然后，使用这些估计值 $\hat{Z}_{ij}$，可以从方程 3.5a 和方程 3.5b 中计算出预测到的概率 $\hat{p}_{ij}$，$i = 1, \cdots, N$，$j = 1, \cdots, M$。注意对每个人来说，对所有的结果估计出来的概

率必须总和为1，$\sum_{j=1}^{M} \hat{p}_{ij} = 1$。多类别logit模型的一个特点就是，对每个结果估计出来的个人概率的平均数（$\bar{p}_j = \sum_{i=1}^{N} \hat{p}_{ij}$，$j = 1, \cdots, M$），等于实际观察到的在这个结果类别中的人所占的比例。

另一种从多类别logit模型中预测概率的方法是对每个结果$j = 1, \cdots, M$，对所有人计算Z_{ij}的均值，如下：

$$\begin{aligned}\bar{Z}_j &= \sum_{i=1}^{N} Z_{ij}/N = \sum_{i=1}^{N}(\sum_{r=1}^{R} \beta_{jr} X_{ir})/N \\ &= \sum_{r=1}^{R} \beta_{jr}(\sum_{i=1}^{N} X_{ir}/N) = \sum_{r=1}^{R} \beta_{jr} \bar{X}_r \end{aligned} \qquad [3.11]$$

然后计算估计到的概率如下：

$$\hat{p}_1 = \frac{1}{1 + \sum_{j=2}^{M} \exp(\bar{Z}_j)} \qquad [3.12a]$$

$$\hat{p}_m = \frac{\exp(\bar{Z}_j)}{1 + \sum_{j=2}^{M} \exp(\bar{Z}_j)} \quad m = 2, \cdots, M \qquad [3.12b]$$

通常，$\bar{p}_1 \neq \hat{p}_1$，$\bar{p}_2 \neq \hat{p}_2$，$\bar{p}_3 \neq \hat{p}_3$，这并不奇怪，因为虽然$\bar{p}_j$和$\hat{p}_j (j = 1, 2, 3)$两者都旨在测量不同结果的整体概率，但是它们的计算方法非常不同。$\bar{p}_j$计算为估计到的个人概率$\hat{p}_{ij}$的均值，而$\hat{p}_{ij}$是从个人决定变量的值中计算出来的。而另一方面，$\hat{p}_j$绕过了个人的概率，而用个人Z_{ij}的平均值$\bar{Z}_j$计算（或者同等地，用决定变量X_{ir}的均值$\bar{X}_r$计算）。

边际效应

决定变量的值的一个小变化对不同结果概率的边际效

应这个问题可以用两种不同的方法表达。X_{ir} 的一个小变化(对个人 i 第 r 个决定变量的值),对某个结果 m 在如下两点上会带来什么变化呢?

1. 概率 $\Pr(Y_i = m)$

2. 风险比 $\Pr(Y_i = m)/\Pr(Y_i = 1)$

第二个问题相对来说比较容易回答,但是第一个问题则难得多。先从简单的问题开始,由方程 3.7,

$$\frac{\partial}{\partial X_{ir}}\log\left(\frac{\Pr\mathrm{ob}(Y_i = m)}{\Pr\mathrm{ob}(Y_i = 1)}\right) = \beta_{mr}$$

因此,对 X_{ir} 的一个小变化,风险比的变化方向可以从相关系数的符号中推断出来;如果 $\beta_{mr} > 0$ 则 $j = m$ 的相对概率增加,如果 $\beta_{mr} < 0$ 则减少。

但是,对 X_{ir} 的一个小变化,观察到结果 $j = m$ 的概率 $\Pr(Y_i = m)$ 其方向的变化,无法从 β_{mr} 的方向上推断出来。原因是,在多类别模型中,某个特定个人某变量值的变化会影响他/她每一种结果的概率。因为这些概率被限定总和为 1,$\Pr(Y_i = m)$ 上升还是下降取决于其他概率的变化情况。所以实际上它不仅取决于 β_{mr} 的方向,还取决于该系数相对于与该变量相关的其他系数的大小情况,即相对于 $\beta_{jr}\, j = 1, \cdots, M,\ j \neq m$。因此,$\partial \Pr\mathrm{ob}(Y_i = m)/\partial X_{ir}$ 不一定与 β_{mr} 的方向相同。

在多类别模型中,确定某变量值的变化对结果概率的影响的最有效的方法是,在其他变量的值保持不变的情况下,比较变化前与变化后的概率。当分析的决定变量是一个虚拟变量时,这种方法最有效。如果第 r 个变量是个虚拟变量,因此 $X_{ir} = 1$ 或者 $X_{ir} = 0$,那么首先假定对所有人 $X_{ir} =$

1，并用方程3.4对每个结果$j=1, \cdots, M$估计Z_{ij}，把这个估计值称为$\hat{Z}_{ij}^{1}$。然后，其他条件相同假定对所有人$X_{ir}=0$，对每个结果$j=1, \cdots, M$估计Z_{ij}，把这个估计值称为$\hat{Z}_{ij}^{0}$，注意$\hat{Z}_{ij}^{1}=\hat{Z}_{ij}^{0}+\hat{\beta}_{jr}$。

用方程3.5a和方程3.5b计算概率的预测值，首先用$Z_{ij}=\hat{Z}_{ij}^{1}$(计算出$\hat{p}_{ij}^{1}$)，然后用$Z_{ij}=\hat{Z}_{ij}^{0}$(计算出$\hat{p}_{ij}^{0}$)。对所有人计算平均概率，$\bar{p}_{j}^{1}=\sum_{i=1}^{N}\hat{p}_{ij}^{1}$和$\bar{p}_{j}^{0}=\sum_{i=1}^{N}\hat{p}_{ij}^{0}$。$\bar{p}_{j}^{1}$与$\bar{p}_{j}^{0}$之间的差就是对所有人，第$r$个变量的值的变化对观察到结果$j(j=1, \cdots, M)$的概率的“平均”作用。

在分析某变量的值变化时，另外一种保持其他变量值不变的方法是把它们的值设为均值，$\bar{X}_{r}=\sum_{i=1}^{N}X_{ir}/N$，然后定义：

$$\bar{Z}_{j}^{1}=\beta_{jr}+\sum_{s\neq r}\beta_{js}\bar{X}_{s} \text{ 和 } \bar{Z}_{j}^{0}=\sum_{s\neq r}\beta_{js}\bar{X}_{s} \qquad [3.13]$$

然后用方程3.5a和方程3.5b计算概率的估计值，先用$Z_{ij}=\bar{Z}_{j}^{1}$(计算出$\hat{p}_{j}^{1}$)，然后用$Z_{ij}=\bar{Z}_{j}^{0}$(计算出$\hat{p}_{j}^{0}$)。$\hat{p}_{j}^{1}$与$\hat{p}_{j}^{0}$之间的差就是对“平均个体”(定义为对所有变量，其值都为平均值的个人)，第r个变量的值的变化对这个人来说，观察到结果$j(j=1, \cdots, M)$的概率的作用。

某个决定变量其值的变化对风险比的影响要简单得多。从方程3.7，$X_{ir}=1$和$X_{ir}=0$的风险比对数的差别，为$\hat{Z}_{ij}^{1}-\hat{Z}_{ij}^{0}=\hat{\beta}_{jr}$，所以对应的风险比的差别为$\exp(\hat{\beta}_{jr})$。这样，某个系数的指数值表示的就是对某个决定变量一个单位的变化(对该结果)所带来的风险比的变化。当然，对虚拟变量来说，考虑一个单位的变化是最合适的。

第5节 | 应用:职业获得

方程设定

作者参考的是从英国1991普查中抽取出来的,关于25—45岁全职男性雇员的职业分类(以及其他特征)的信息。[37]职业的分类为非技术/半技术工人(UNS)、有技术体力/非体力劳动者(SKL)和专业/管理/技术人员(PMT)。

表3.1 男性全职雇员的样本统计值:按种族类别划分*

	白人	黑人	印第安人
样本量	96297	863	1572
年龄(年)	35.4	32.6	34.9
%职业类别			
专业/管理/技术人员(PMT)	41.7	29.3	35.1
有技术体力/非体力劳动者(SKL)	40.7	50.2	39.8
非技术/半技术工人(UNS)	17.6	20.5	25.2
%具18岁或以上的文凭			
学位文凭	15.1	7.9	17.4
副学位文凭	9.2	7.1	6.6
没有18岁或以上的文凭	75.7	85.0	76.0
%出生在			
英国	96.2	54.6	7.6
国外	3.8	45.4	92.4
%居住在			
伦敦	10.1	51.5	43.6
北部	50.3	24.9	30.2
南部	39.6	23.6	26.2

注:* 信息来自英国1991年人口普查。

在这 98732 个人当中，有 96297 个"白人"，863 个"黑人"和 1572 个印第安人。表 3.1 显示了样本统计量的主要特征。这张表非常清楚地说明，职业地位在白人、黑人和印第安人之间存在着差别。比如说，42%的白种男性雇员的工作属于 PMT 类别，但只有 35%的印第安人，以及 29%的黑人从事类似的工作。这些差别并不能被特征的差别解释掉：与 15%的白种工人相比，17%的印第安雇员有学位文凭。

依据前面阐述的选择构成，看上去白人、黑人和印第安人在做选择时面临不同系列的限制。这可能是因为，与具有同等资格的白人相比，处于少数民族组别中的个人在劳动力市场上面临着劣势。也可能是因为与白人相比，处于少数民族组别中的个人具有更不利的劳工特征（此后称之为特征）。因此，职业表现的组间差别可能同时是种族与特征不利因素的结果。

关键问题是：这些差别中有多少是种族不利因素的结果，又有多少归因于特征的不利因素？前一部分所描述的多类别 logit 方法可以用来回答这个问题。

起点是对样本中每个 $i = 1, \cdots, 98732$ 男性定义因变量 Y_i 以使：

$Y_i = 1$ 如果这个人受雇于 UNS 职业类别；

$Y_i = 2$ 如果这个人受雇于 SKL 职业类别；

$Y_i = 3$ 如果这个人受雇于 PMT 职业类别。

在这个多类别模型设定中使用的决定变量包括：

年龄

AGE 以年计算：通过设定 25 岁的人 AGE = 0 进行标准化。

教育[38]

HED = 1,如果这个人有学位程度文凭,否则 HED = 0;

MED = 1,如果这个人有高于 A 程度文凭(post-A level),但低于学位文凭;否则 MED = 0。

学习领域[39]

SCI = 1,如果为自然科学相关领域;否则 SCI = 0;

BUS = 1,如果为商学相关领域;否则 BUS = 0。

种族[40]

BLK = 1,如果这个人的种族为加勒比黑人,否则 $BLK = 0$;

$IND = 1$,如果这个人的种族为印第安人;否则 IND = 0。

出生国家

$OVB = 1$,如果这个人在英国以外出生;否则 $OVB = 0$。

居住地区[41]

$STH = 1$,如果住在英国南部[42](伦敦除外),否则 $STH = 0$;

NTH = 1,如果住在英国北部;[43]否则 NTH = 0。

以这些变量为基础,那两个风险比(对数)方程(方程 3.7,对 $j = 2, 3$)确定如下:

$$\begin{aligned}\log\left(\frac{\Pr(Y_i = j)}{\Pr(Y_i = 1)}\right) = {} & \alpha_{j0} + \beta_{j0} * BLK_i + \gamma_{j0} * IND_i + \alpha_{j1} * OVB_i \\ & + \beta_{j1} * BLK_i * OVB_i + \gamma_{j1} * IND_i * OVB_i \\ & + \alpha_{j2} * HED_i + \beta_{j2} * BLK_i * HED_i \\ & + \gamma_{j2} * IND_i * HED_i + \alpha_{j3} * MED_i \\ & + \beta_{j3} * BLK_i * MED_i + \gamma_{j3} * IND_i * MED_i\end{aligned}$$

$$
\begin{aligned}
&+\alpha_{j4} * NTH_i + \beta_{j4} * BLK_i * NTH_i \\
&+\gamma_{j4} * IND_i * NTH_i + \alpha_{j5} * STH_i \\
&+\beta_{j5} * BLK_i * STH_i + \gamma_{j5} * IND_i * STH_i \\
&+\theta_{j1} * AGE_i + \theta_{j2} * AGE^2 + \theta_{j3} * BUS_i \\
&+\theta_{j4} * SCI_i + \theta_{j5} * BUS_i * HED_i \\
&+\theta_{j6} * SCI_i * HED_i \qquad [3.14]
\end{aligned}
$$

$j=1$ 这个结果——也就是，处于 UNS 这个职业类别——在此后称为“基准”结果。这个结果的系数被设定为等于 0，其他结果的风险比定义为关于这个基准结果概率的风险比。[44]

在方程 3.14 中，α_{jr} 是“白人”的系数，β_{jr} 和 γ_{jr} 分别表示由于是黑人和印第安人，而对这些系数的附加的贡献。另一方面，θ_{jr} 是假定没有“种族维度”的系数，意即它们（被假定）在种族方面没有变化。从方程 3.14，对一个 25 岁（AGE = 0），英国出生的（OVB = 0），白种（BLK = IND = 0）男性雇员，且没有 18 岁或以上的文凭（HED = MED = 0），居住在伦敦（NTH = STH = 0），其风险比对数为 α_{j0}。如果有这么一个人是黑人或者印第安人，那么他的风险比对数会分别变化 β_{j0} 和 γ_{j0}，分别变为：$(\alpha_{j0}+\beta_{j0})$ 和 $(\alpha_{j0}+\gamma_{j0})$。

举例来说，如果 $\beta_{j0}<0$，那么处于职业类别 j 的概率相对处于 UNS 类别的概率的比率，对一个白人来说会比具有同等条件的黑人要高。系数 β_{j0} 和 γ_{j0} 的大小测量了黑人与印第安人相对于白人来说，（在职业类别 j 方面）面临的“种族不利因素”的程度，在如上描述的特征的情况下。[45] 方程 3.14 中的交互项，涉及种族变量 BLK 和 IND，允许种族不利因素的程度随着个人的某些非种族特征而变化。这些非种族特

征为出生国家、居住地区和文凭程度。对具有上述特征的男性来说,处于职业类别 j 的风险比对在国外出生的(OVB = 1)白人来说是 $\alpha_{j0}+\alpha_{j1}$,对在国外出生的(OVB = 1)黑人来说是 $\alpha_{j0}+\alpha_{j1}+\beta_{j0}+\beta_{j1}$。

方程 3.14 包含的非种族特征交互项中,其中有一系列是学习领域和文凭程度之间的交互作用。这些交互项的相关系数为 θ_{j5} 和 θ_{j6},且其值不等于 0。这意味着,在决定风险比的值时,不仅学习领域和文凭程度在分开考虑时有重要影响;而且学科与文凭是如何融合而产生作用也很重要,如一个具有自然科学相关学位的毕业生。

方程统计量

对方程 3.14 的估计,首先对它的系数没有附加任何限制,然后限定它的某些系数为 0。方程 3.14 一共包括 48 个系数,在 SKL(j = 1)和 PMT(j = 2)这两个方程中分别各占 24 个。在这 48 个系数中,有 13 个被设为 0。这些设为 0 的系数,各自并不显著地不等于 0,而且,似然比检验 $\chi^2(13)=8.1$,并没有拒绝它们全部等于 0 的联合假设。表 3.2 和表 3.3 分别表示,在没有附加任何限制和附加了零限制的情况下,对方程 3.14 的估计结果。对这些没有限制和有零限制的结果的比较显示,附加零限制并没有在本质上影响(quanl-itatively affect)那些没有被设为 0 的系数的估计值。表 3.2 和表 3.3 中的 z-ratios 是估计系数相对于它们的估计标准误差的比率。在相关系数等于 0 的零假设下,这些 z-ratios 渐进地服从 $N(0, 1)$分布。[46]

表 3.2 职业选择的多类别 logit 估计结果:完整模型

Multinomial Regression Log Likelihood = −86948.078; *Number of obs* = 98732; $LR\chi^2(46)$ = 30880.36; $Prob > \chi^2$ = 0.0000; *Pseudo* R^2 = 0.1508

y	*Coefficient*	*Standard Error*	*z*	$P > \|z\|$	[95% *Conf. Interval*]	
$y=2$						
age	0.0261899	0.0064884	4.036	0.000	0.0134729	0.0389069
age2	−0.0009772	0.000306	−3.194	0.001	−0.0015768	−0.0003775
blkcb	0.3655553	0.1672964	2.185	0.029	0.0376603	0.6934503
blkovs	0.08195	0.1887671	0.434	0.664	−0.2880266	0.4519266
ind	0.4707812	0.2790797	1.687	0.092	−0.0762051	1.017767
indovs	−0.6064709	0.2728072	−2.223	0.026	−1.141163	−0.0717786
ovsbn	−0.109104	0.0540097	−2.020	0.043	−0.214961	−0.003247
north	−0.1445068	0.0355612	−4.064	0.000	−0.2142055	−0.0748082
indnth	−0.4185296	0.1586727	−2.638	0.008	−0.7295225	−0.1075368
blknth	−0.7146398	0.2170033	−3.293	0.001	−1.139959	−0.2893211
south	−0.0776431	0.0365635	−2.124	0.034	−0.1493062	−0.0059799
indsth	−0.3459856	0.1654501	−2.091	0.037	−0.6702619	−0.0217094
blksth	−0.6664942	0.225652	−2.954	0.003	−1.108764	−0.2242244
highed	1.083631	0.1822909	5.945	0.000	0.7263477	1.440915
indhe	−0.2140724	0.4490211	−0.477	0.634	−1.094138	0.6659928
blkhe	−1.26384	0.7230326	−1.748	0.080	−2.680958	0.1532775
mided	0.6060102	0.2709862	2.236	0.025	0.0748871	1.137133
indme	−0.3332887	0.4428193	−0.753	0.452	−1.201199	0.5346213
blkme	0.9548953	1.034856	0.923	0.356	−1.073386	2.983177
subbus	1.205355	0.3189165	3.780	0.000	0.5802901	1.83042
subsci	0.5103959	0.2820694	1.809	0.070	−0.0424499	1.063242
bush	−0.6332213	0.4138681	−1.530	0.126	−1.444388	0.1779454
scih	−0.795496	0.3545109	−2.244	0.025	−1.490325	−0.1006674
_cons	0.7755227	0.042255	18.353	0.000	0.6927043	0.858341
$y=3$						
age	0.0893431	0.0073103	12.222	0.000	0.0750152	0.103671
age2	−0.0024609	0.0003397	−7.244	0.000	−0.0031266	−0.0017951
blkcb	−0.0672642	0.1950314	−0.345	0.730	−0.4495187	0.3149904
blkovs	−0.5708287	0.2236755	−2.552	0.011	−1.009225	−0.1324329
ind	0.0918407	0.3348558	0.274	0.784	−0.5644646	0.748146
indovs	−0.7893394	0.3292557	−2.397	0.017	−1.434669	−0.1440101
ovsbn	0.0932756	0.0565007	1.651	0.099	−0.0174637	0.204015

续表

Multinomial Regression Log Likelihood = −86948.078; *Number of obs* = 98732; *LR* χ^2(46) = 30880.36; *Prob* > χ^2 = 0.0000; *Pseudo* R^2 = 0.1508						
y	*Coefficient*	*Standard Error*	*z*	*P* > |*z*|	[95% *Conf. Interval*]	
north	−0.6113055	0.037637	−16.242	0.000	−0.6850727	−0.5375384
indnth	−0.4696822	0.1943706	−2.416	0.016	−0.8506416	−0.0887228
blknth	−0.1352359	0.2598729	−0.520	0.603	−0.6445774	0.3741057
south	−0.1697707	0.0383309	−4.429	0.000	−0.244898	−0.0946434
indsth	−0.5134623	0.1951871	−2.631	0.009	−0.8960221	−0.1309025
blksth	−0.418329	0.268549	−1.558	0.119	−0.9446753	0.1080174
highed	3.877128	0.1713991	22.620	0.000	3.541191	4.213064
indhe	−0.1698475	0.4062421	−0.418	0.676	−0.9660674	0.6263724
blkhe	−1.139974	0.6128429	−1.860	0.063	−2.341124	0.0611758
mided	2.850769	0.2490918	11.445	0.000	2.362558	3.33898
indme	−0.2049551	0.4154452	−0.493	0.622	−1.019213	0.6093026
blkme	0.743211	1.023254	0.726	0.468	−1.262329	2.748751
subbus	0.8151578	0.298235	2.733	0.006	0.230628	1.399688
subsci	0.2500103	0.2599529	0.962	0.336	−0.2594879	0.7595085
bush	−0.0414593	0.3898265	−0.106	0.915	−0.8055052	0.7225865
scih	0.1600076	0.3281777	0.488	0.626	−0.4832088	0.803224
_cons	−0.0164003	0.0462775	−0.354	0.723	−0.1071026	0.0743019

注:因变量 $y=1$ 为对照组。

表 3.3 职业选择的多类别 logit 估计结果:限制模型

Multinomial Regression Log Likelihood = −86952.126; *Number of obs* = 98732; *LR* χ^2(33) = 30872.27; *Prob* > χ^2 = 0.0000; *Pseudo* R^2 = 0.1508						
y	*Coefficient*	*Standard Error*	*z*	*P* > |*z*|	[95% *Conf. Interval*]	
y = 2						
age	0.026323	0.0064795	4.062	0.000	0.0136234	0.0390227
age2	−0.0009813	0.0003057	−3.210	0.001	−0.0015805	−0.0003822
blk	0.426683	0.1116093	3.823	0.000	0.2079328	0.6454332
blkovs	(dropped)					
ind	0.4144723	0.2152989	1.925	0.054	−0.0075059	0.8364504
indovs	−0.565894	0.2215624	−2.554	0.011	−1.000148	−0.1316396
ovsbn	−0.106519	0.05151	−2.068	0.039	−0.2074767	−0.0055613

续表

Multinomial Regression Log Likelihood = −86948.078; *Number of obs* = 98732; $LR\chi^2(46)$ = 30880.36; *Prob* > χ^2 = 0.0000; *Pseudo* R^2 = 0.1508						
y	*Coefficient*	*Standard Error*	*z*	*P* > |*z*|	[95% *Conf. Interval*]	
north	−0.1438706	0.0351996	−4.087	0.000	−0.2128606	−0.0748806
indnth	−0.4152823	0.1546391	−2.685	0.007	−0.7183694	−0.1121952
blknth	−0.6568392	0.1761849	−3.728	0.000	−1.002155	−0.3115232
south	−0.0765746	0.0362333	−2.113	0.035	−0.1475906	−0.0055585
indsth	−0.3417737	0.1632409	−2.094	0.036	−0.66172	−0.0218275
blksth	−0.6677784	0.2086548	−3.200	0.001	−1.076734	−0.2588226
highed	1.373433	0.1088435	12.618	0.000	1.160104	1.586763
indhe	(dropped)					
blkhe	−1.312484	0.7222905	−1.817	0.069	−2.728147	0.1031797
mided	0.8347366	0.1476385	5.654	0.000	0.5453704	1.124103
indme	(dropped)					
blkme	(dropped)					
subbus	0.9348005	0.1923779	4.859	0.000	0.5577467	1.311854
subsci	0.2707989	0.1355025	1.998	0.046	0.0052189	0.5363788
bush	−0.6003602	0.1675312	−3.584	0.000	−0.9287153	−0.272005
scih	−0.9608239	0.1601125	−6.001	0.000	−1.274639	−0.6470093
_cons	0.7737729	0.041826	18.500	0.000	0.6917954	0.8557504
y = 3						
age	0.0895118	0.0073026	12.258	0.000	0.0751989	0.1038247
age2	−0.0024658	0.0003395	−7.263	0.000	−0.0031313	−0.0018004
blk	(dropped)					
blkovs	−0.6781675	0.154667	−4.385	0.000	−0.9813092	−0.3750257
ind	(dropped)					
indovs	−0.7063556	0.1295174	−5.454	0.000	−0.9602051	−0.4525062
ovsbn	0.0954136	0.0553704	1.723	0.085	−0.0131104	0.2039377
north	−0.6105677	0.0368792	−16.556	0.000	−0.6828497	−0.5382858
indnth	−0.4542501	0.181704	−2.500	0.012	−0.8103834	−0.0981168
blknth	(dropped)					
south	−0.1682364	0.0376312	−4.471	0.000	−0.2419921	−0.0944807
indsth	−0.501347	0.1887684	−2.656	0.008	−0.8713262	−0.1313678
blksth	−0.4365967	0.2228379	−1.959	0.050	−0.8733511	0.0001576
highed	4.171942	0.0844821	49.383	0.000	4.00636	4.337524

续表

Multinomial Regression Log Likelihood $= -86948.078$; Number of obs $= 98732$; LR $\chi^2(46) = 30880.36$; Prob $> \chi^2 = 0.0000$; Pseudo $R^2 = 0.1508$						
y	*Coefficient*	*Standard Error*	*z*	$P>\|z\|$	[95% *Conf. Interval*]	
indhe	(dropped)					
blkhe	−1.198389	0.6072452	−1.973	0.048	−2.388567	−0.0082099
mided	3.089609	0.0695418	44.428	0.000	2.95331	3.225909
indme	(dropped)					
blkme	(dropped)					
subbus	0.5326858	0.1343444	3.965	0.000	0.2693757	0.795996
subsci	(dropped)					
bush	(dropped)					
scih	(dropped)					
_cons	−0.0188423	0.0454545	−0.415	0.678	−0.1079314	0.0702469

注:因变量 $y = 1$ 为对照组。
Mlogit: likelihood-ratio test $\chi^2(13)=8.1$; Prob$>\chi^2=0.8372$。

估计值

正如前面的讨论所强调的,表 3.2 和表 3.3 中系数估计值的方向反映了在其他条件同等的情况下,对应于该系数相关变量值的变化,风险比 $\Pr(Y_i = j)/\Pr(Y_i = 1)$ 的变化方向。它并不反映单个概率 $\Pr(Y_i = j)$的变化方向。在这一部分中报告的估计结果是表 3.3 中的约束系数。这是因为,当把完整的方程设定(见表 3.2)应用到数据中时,我们发现这些变量只有一小部分对风险比有显著的影响。在大多数情况下,被排除在外的变量是那些涉及某个非种族变量 X 与种族变量 BLK 和/或 IND 的交互项。这意味着当变量 X 本身对个人 i 属于某个特定职业类别的风险比有显著影响时,这个人属于哪个种族并不会改变这种影响。

表3.3所示的估计结果，点出了三种对改善个人处于SKL或PMT职业类别风险比来说很重要的特征：[47]

住在伦敦；

具有18岁或18岁以上的文凭，如果取得学位或学习与商业相关的科目会更好；

出生在英国。

住在伦敦有两点好处：首先，对所有人有一个整体的益处。这来源于如下情况：居住在北部或者南部的男性，在其他条件相同时，与居住在伦敦的人相比，其处于SKL或PMT职业类别的风险比更低[48]（方程3.14中 $\hat{\alpha}_{j4}$，$\hat{\alpha}_{j5}<0$，$j=2$，3）。其次，对黑人和印第安人来说还有特定的好处。因为住在伦敦而不是在伦敦之外，黑人和印第安人与白人相比，在处于SKL或PMT职业类别的风险比上具有更大的提升。相比于住在伦敦，住在伦敦之外的人其处于SKL或PMT职业类别的风险比的下降程度，对黑人[49]或者印第安人比对白人更大（$\hat{\beta}_{j4}$，$\hat{\beta}_{j5}$，$\hat{\gamma}_{j4}$，$\hat{\gamma}_{j5}<0$，$j=2$ 以及 $\hat{\beta}_{j5}$，$\hat{\gamma}_{j4}$，$\hat{\gamma}_{j5}<0$，$j=3$）。

从这种意义上说，相对于白人，伦敦比英国其他地区对印第安人和黑人更加友善。特别是住在伦敦的印第安人（$NTH_i=STH_i=0$），并不比白人在处于PMT职业类别的风险比方面具有劣势，因为 $\hat{\gamma}_{j0}=0$，$j=3$。住在伦敦之外（$NTH_i=1$）对所有人这个风险比都下降，但是对印第安人来说下降程度要比对白人更大。从这种意义上说，在PMT职业类别中，印第安人在伦敦相对于白人所享受的种族同等情况，在伦敦以外的地区减弱了。“伦敦效应”对黑人最有利。在全职男性雇员中，只有10%的白人住在伦敦，但有44%的印第安人和53%的黑人住在这个地区。

出生在英国之外一直是一个劣势:处于 SKL 或 PMT 职业类别的风险比对英国出生的人比在国外出生的人更大(方程 3.14 中 $\hat{\alpha}_{j1} < 0, j = 2, 3$)。但是,因为出生地这个变量与种族的交互项是负的($\hat{\beta}_{j1}\hat{\gamma}_{j1} < 0, j = 2, 3$),出生在国外的劣势对印第安人和黑人来说,要比白人更大,即由于出生在国外而导致的风险比的降低程度对印第安人和黑人比白人更大。

拥有 18 岁或 18 岁以上的文凭,不管是副学位还是学位(或更高)文凭,两者都提高了处于 SKL 或 PMT 职业的风险比,其中学位文凭的作用比副学位文凭的作用更大。

除了文凭之外,所取得文凭的科目也有重要作用。商业类科目的文凭为进入 SKL 和 PMT 这两种职业都提供了最好的途径。计量经济学的结果是对男性印第安人来说,方程 3.14 中的 γ_{j2}(HED)和 γ_{j3}(MED)在 SKL($j = 1$)和 PMT($j = 2$)这两种职业中都显著地等于 0。对男性黑人来说,在 PMT 职业类别中 β_{j2}(HED)明显是负的。男性黑人相比印第安人或白人来说,具有更低的学位文凭回报率,但是印第安人和白人的回报率相同。

相关文献指出,那些在自己的国家获得文凭的移民,尤其是从发展中国家来的移民,持有这样一种观念(不管是否有理):这些文凭与在接收国家获得的同等文凭相比,更"没有价值"。遗憾的是,我们的数据并没有对那些出生在国外的人,记录他们迁入英国的时间。因此,虽然我们知道样本中大部分黑人和印第安人出生在国外(表 3.1),但是却没有关于他们到达英国时的年龄方面的信息,这意味着,我们无法推断他们在哪里获得他们的 18 岁或 18 岁以上的文凭(如

果有的话)。

概率估计值

利用估计出来的 $\hat{Z}_{ij}$——记住 $\hat{Z}_{ij}=\sum_{r=1}^{R}\hat{\beta}_{jr}\times X_{ir}$ 是用表 3.3 中的估计值和每个人的决定变量的值联合估计出来的——STATA 会通过计算方程3.5a中的 $\hat{p}_{i1}$ 和 3.5b 中的 $\hat{p}_{i2}$ 和 $\hat{p}_{i3}$,对样本 98732 个人中的每个人估计处于三种不同职业类别的概率。对所有人,以及对白人、黑人和印第安人这三组人,分别计算出来的这些个人概率的均值,分别记为 $\bar{p}_j$, $\bar{p}_{Wj}$, $\bar{p}_{Bj}$ 和 $\bar{p}_{Ij}$ $(j=1, 2, 3)$。如表 3.4 所示,这些平均概率(估计到

表 3.4 白人、黑人和印第安人处于不同职业类别的概率估计值*

	概率估计值		
	UNS	SKL	PMT
个人概率估计值的均值			
总体	17.7	40.8	41.5
白人	17.5	40.7	41.8
黑人	20.5	50.2	29.3
印第安人	25.1	39.8	35.1
用决定变量的均值计算而得的概率估计值			
总体	14.3	41.8	43.9
白人	14.2	41.7	44.1
黑人	19.6	52.0	28.4
印第安人	22.0	43.3	34.7

注:* 用表 3.3 的多类别 logit 估计结果计算而得。
UNS = 非技术/半技术工人。
SKL = 有技术体力/非体力劳动者。
PMT = 专业/管理/技术人员。

的个人概率的均值)表示,比如,对所有男性雇员和白人,黑人以及印第安男性雇员估计到的处于 PMT 职业类别的平均概率分别为 41.5%,41.8%,29.3%和 35.1%。应该强调的是,对所有人和对每个种族组别内的个人所估计到的处于三种职业类别的平均概率,正好等于所对应的每个分类的样本比例。[50]

这是多类别 logit 模型的一个特点,估计到的个人结果概率的均值总是等于这些结果的样本比例。

估计概率的另一种方法是对所有人,以及对白人、黑人和印第安人这三组人计算 $\hat{Z}_{ij}$ 的均值。如果这些均值分别记为 $\bar{Z}_j$,$\bar{Z}_{Wj}$,$\bar{Z}_{Bj}$,$\bar{Z}_{Ij}$,那么方程3.12a就可以用来对所有人和每一组人计算第一种结果的概率 $\hat{P}_1$,$\hat{P}_{W1}$,$\hat{P}_{B1}$和 $\hat{P}_{I1}$),方程 3.12b 就可以用来对所有人和每一组人计算第二种和第三种结果的概率 $\hat{P}_j$,$\hat{P}_{Wj}$,$\hat{P}_{Bj}$和 $\hat{P}_{Ij}$($j=2, 3$)。[51]这些概率,如表 3.4 所示(用决定变量的均值估计的概率),显示在决定变量为均值时,所有人、白人、黑人和印第安男性雇员处于 PMT 职业的概率估计值分别为 43.9%,44.1%,28.4%和 34.7%。

这些概率估计值与该表上半部分报告的概率不一样,这并不奇怪。举例来说,虽然 $\bar{p}_{Wj}$和 $\hat{P}_{Wj}$($j=1, 2, 3$)两者都旨在测量白人处于不同职业类别的总体概率,但是他们的计算方法非常不同。$\bar{p}_{Wj}$是(对 96297 个白人)估计到的个人概率 $\hat{p}_{ij}$的均值,而 $\hat{p}_{ij}$是用个人决定变量的值计算出来的。另一方面,$\hat{P}_{Wj}$——绕过了个人概率——反而是用样本 96297 个白人 Z_{ij}的均值 $\bar{Z}_{Wj}$计算出来的。[52]因此,总的来说,对任何组别 g,$\bar{P}_{g1} \neq \hat{P}_{g1}$;$\bar{P}_{g2} \neq \hat{P}_{g2}$;

$\bar{P}_{g3} \neq \hat{P}_{g3}$。

通过模拟获取边际效应

在之前的讨论中，我们强调了对 X_k 的一个小变化观察到结果 j 的概率 $\Pr(Y_i = j)$ 的变化方向不能从 β_{jr} 的方向中推断出来，因为 $\partial \text{Prob}(Y_i = j)/\partial X_{ir}$ 不一定与 β_{jr} 的方向相同。在多变量 logit 模型的情况下，只有风险比的变化方向能从系数的方向中估计出来。因此，前面对估计结果的讨论就是关于这些风险比的。

但是，可能有人会对概率 $\Pr(Y_i = j)$ 而不是相对于某个基准结果的风险比 $\Pr(Y_i = j)/\Pr(Y_i = 1)$ 感兴趣。在这个应用实例中，可能有人会对处于某个具体职业类别的概率在不同的种族间如何存在差别特别感兴趣。由于系数估计值并不直接对此给出答案，一种可以替代的方法就是通过模拟的窗口观察这些结果。我们可以通过计算在不同的假设情形下不同结果的概率而使模型的结果变得清晰明显。[53] 更具体地，种族的作用可以通过比较当种族虚拟变量取不同值时的概率分析出来，在做这些比较时其他变量的值保持不变。

假设样本中所有 98732 个人都是白人，即对所有 $i = 1, \cdots, N$, $IND_i = BLK_i = 0$。让 P_{ij}^{W} 表示在这个假设情形下个人 i 处于职业类别 $j (j = 1, 2, 3)$ 的概率。现在假设样本中所有的 98732 个人都是黑人，即对所有 $i = 1, \cdots, N$, $BLK_i = 1$，让 P_{ij}^{B} 表示在这个假设情形下个人 i 处于职业类别 $j (j = 1, 2, 3)$ 的概率。[54] 根据方程 3.14 的设定，这两个

概率P_{ij}^W和P_{ij}^B是当样本中所有人的特征分别用白人(α_{jr})和黑人的系数($\alpha_{jr}+\beta_{jr}$)进行评估时的结果。此后称它们为"种族"概率,因为它们是用只在种族特征上不同的个人所组成的样本计算出来的。当应用到方程3.5a和方程3.5b时,系数估计值和$\hat{\alpha}_{jr}$和$\hat{\alpha}_{jr}+\hat{\beta}_{jr}$产生种族概率的估计值$\hat{P}_{ij}^W$和$\hat{P}_{ij}^B$。

如果$\bar{p}_j^W=\sum_{i=1}^N \hat{P}_{ij}^W/N$, $\bar{p}_j^B=\sum_{i=1}^N \hat{P}_{ij}^B/N$是个人概率估计值各自的均值(其中后者是在两组假定的情形下计算出来的)。那么,种族概率的均值$\bar{p}_j^W$和$\bar{p}_j^B$之差就完全是应用于(样本中N个人)给定特征时,不同系列的系数(白人和黑人)之间差别的结果。因此,这些差异可以完全归因于对这两组人不平等对待的结果,而这两组人除了种族不同,在其他每个方面都完全相同。更简明地说,可以归因于黑人遭受"种族劣势"这个事实。类似的应用也可以在假定样本中每个人都为印第安人的情况下实行。

这些种族概率的估计值,分别对白人、黑人和印第安人这三组人,见表3.5上面的那部分。这部分显示如果样本中每个人的特征都用黑人的系数$\alpha_{jr}+\beta_{jr}$,而不是用与他们相关的种族系数估计,[55]那么总样本中40%的人会在PMT职业类别中,39%会在SKL,21%会在UNS职业中。

另外一种用这个方式计算种族概率的方法是,比较当种族变量取不同的值但其他变量都在各自情形下取样本均值时的概率。用$\bar{Z}_j^W$和$\bar{Z}_j^B$表示当样本中所有人都(被视为)是白人和样本中所有人都(被视为)是黑人时$\hat{Z}_{ij}=$

$\sum_{r=1}^{R} \hat{\beta}_{jr} X_{ir}$ 的均值。[56]

那么白人和黑人处于不同职业类别中的概率(分别用 $\hat{P}_j^W$ 和 $\hat{P}_j^B$, $j=1, 2, 3$ 表示)是用 $\bar{Z}_j^W$ 和 $\bar{Z}_j^B$ 计算出来的，对第一种职业类别用方程 3.12a,对第二和第三种用方程 3.12b。这些估计出来的概率的值,分别对白人、黑人和印第安人列于表 3.5。当概率是用样本的平均特征计算时，估计到更大比例的白人和黑人(更小比例的印第安人)处于 PMT 类别($\hat{P}_3^W = 44.2\%$, $\hat{P}_3^B = 41.0\%$, $\hat{P}_3^I = 32.6\%$),相对于当概率是用个人概率的均值计算时($\bar{P}_3^W = 41.8\%$, $\bar{P}_3^B = 39.8\%$, $\bar{P}_3^I = 33.7\%$)。这是因为印第安人具有比均值更高的高等教育文凭特征(见表 3.1),而这种特征对获得 PMT 职业来说是成功的主要决定因素。因此所有人都降低到样本均值时,印第安人会“损失”而其他人则“获利”。

但是,正如表 3.5(样本比例)所示,只有 29%的黑人男性雇员实际上在 PMT 职业内。黑人和白人在职业类别 j 的样本比例,分别用 s_j^B 和 s_j^W 表示,通常不同于种族概率,$\bar{P}_j^B$ 和 $\bar{P}_j^W$(以及 $\hat{P}_j^B$ 和 $\hat{P}_j^W$)。这反映了一个情况,即黑人和白人不仅在他们是如何被对待的这方面不同,而且在他们的特征方面也不同。黑人处于 PMT 职业的样本比例($s_3^B = 29.3\%$)比估计到的处于该职业的种族概率($\bar{P}_3^B = 39.8\%$ 或者 $\hat{P}_3^B = 41.0\%$)低的情况是因为,相对于整体样本,黑人的特征更不适合 PMT 职业这个实际情况。或者说,黑人相对于白人,在 PMT 职业方面会同时遇到种族和特征方面的劣势。这些劣势之和被称为总体劣势。

表 3.5 白人、黑人和印第安人处于不同职业类别的“种族”概率估计值*

	概率估计值		
	UNS	SKL	PMT
个人概率估计值的均值			
白人	17.5	40.7	41.8
黑人	20.7	39.5	39.8
印第安人	19.3	47.0	33.7
样本比例			
白人	17.5	40.7	41.8
黑人	20.5	50.2	29.3
印第安人	25.1	39.8	35.1
用决定变量的均值计算而得的概率估计值			
白人	14.2	41.6	44.2
黑人	19.3	39.7	41.0
印第安人	16.6	50.8	32.6

注:* 用表 3.3 的多类别 logit 估计结果计算而得。
UNS = 非技术/半技术工人。
SKL = 有技术体力/非体力劳动者。
PMT = 专业/管理/技术人员。

测量职业劣势

黑人同白人相比,在职业类别 j 中遇到的平均的种族劣势的一个测量指标是 λ_j^B,其中:

$$\lambda_j^B = \bar{p}_j^B / \bar{p}_j^W \qquad [3.15]$$

如果这两个概率相等,则 $\lambda_j^B = 1$,没有种族劣势。但是,如果 $\bar{p}_j^B < \bar{p}_j^W$,则 $\lambda_j^B < 1$,对职业类别 j 存在黑人种族劣势[57],λ_j^B 离 1 越远,则这种劣势的规模越大。

黑人同白人相比,在职业类别 j 中遇到的平均的总体劣势的一个测量指标是 μ_j^B,其中:

$$\mu_j^B = s_j^B / s_j^W \qquad [3.16]$$

如果这两个样本比例相等,则 $\mu_j^B = 1$,没有总体劣势。但是如果 $s_j^B < s_j^W$,则 $\mu_j^B < 1$,对职业类别 j 存在黑人总体劣势[58],μ_j^B 离 1 越远,则这种劣势的规模越大。

黑人同白人相比,在职业类别 j 中遇到的平均特征劣势的一个测量指标是 δ_j^B,其中 δ_j^B 是总体劣势与种族劣势的比:$\delta_j^B = \mu_j^B / \lambda_j^B$。

如果 $\delta_j^B = 1$,则黑人没有面临特征劣势,因为 $\mu_j^B = \lambda_j^B$。这种情况下,黑人与白人样本比例的比等于对应的种族概率的比。如果种族概率的比小于 1,则完全归因于黑人与白人相同的特征被区别对待,对后者的评估比对前者更加有利。但是,如果 $\delta_j^B < 1$,则存在特征劣势,而且 δ_j^B 离 1 越远这种劣势的规模越大。黑人在他们的特征方面受到惩罚 $\mu_j^B < \lambda_j^B$。他们因为具有较差的特征而吃亏,此外这些特征还(可能)更加不利地被评估,相比假若它们是白人的特征的话。最后,如果 $\delta_j^B > 1$,则黑人享受了特征优势 $\mu_j^B > \lambda_j^B$。黑人因为具有更好的特征而弥补了可能的不公平对待所带来的损失。其实,如果 δ_j^B 的值足够大那么就可能出现 $\mu_j > 1$,即使 $\lambda_j < 1$。换句话说,黑人较有优势的特征可以甚于中和他们可能遇到的任何不公平对待。表 3.6 展示了三个不同种族的这三种不同的劣势估计值。

表 3.6 中“种族劣势”这一栏下面的数字表示,如果印第安人和白人配以一组共同的特征(即总样本的特征)的话,那么印第安人处于 PMT 类别的概率等于对应白人概率的 81%。黑人的类似经历会产生 95%这个数值。但实际上,印

第安人和黑人作为不同的组别，具有与总体样本不同的特征，这意味着印第安人和黑人在 PMT 类别的样本比例，分别为对应白人比例的 84%和 70%。印第安人从种族的 81%这个数值上升到样本的 84%这个数，以及黑人对应地从 95%下降到 70%，可能分别归因于，印第安人"具有优势的"特征和黑人"具有劣势的"特征。表 3.6 显示，在 PMT 职业方面，相对于白人，印第安人有 4%的特征优势，但是黑人则有 15%的特征劣势。

表 3.6 黑人和印第安人相对于白人面临的种族、特征和总体劣* 的估计结果

	种族劣势（%）	特征劣势（%）	总体劣势（%）
用个人概率估计值的均值计算而得			
黑人/白人			
$j=1$(SKL)	0.97	1.27	1.23
$j=2$(PMT)	0.95	0.74	0.70
印第安人/白人			
$j=1$(SKL)	1.15	0.85	0.98
$j=2$(PMT)	0.81	1.04	0.84
用决定变量均值计算而得的概率估计值计算而得			
黑人/白人			
$j=1$(SKL)	0.95	1.26	1.23
$j=2$(PMT)	0.93	0.75	0.70
印第安人/白人			
$j=1$(SKL)	1.22	0.80	0.98
$j=2$(PMT)	0.73	1.15	0.84

注：* 若取值 >1，则为具有优势。用表 3.5 中的数字计算而得。
UNS = 非技术/半技术工人。
SKL = 有技术体力/非体力劳动者。
PMT = 专业/管理/技术人员。

在SKL这个类别，印第安人相对于白人具有15%的特征劣势，而黑人相对于白人则有27%的特征优势。

我们也可以把种族劣势定义为 $\hat{P}_j^B/\hat{P}_j^W$——即通过设定除了种族之外的其他变量等于它们的样本均值而获得的概率比（见表3.5的下部分）。特征劣势现在重新计算为总体劣势[59]与这种新的种族劣势估计值之间的比。新的种族和特征劣势的值（以及没有变化的总体劣势的值）见表3.6。

第 6 节 | 条件 logit 模型与不相关选项的独立性

在条件 logit 模型中,结果概率 $\Pr(Y_i = j)$仅仅取决于选择的属性而不取决于做选择的个人的特征。或者说,条件 logit 模型由方程 3.3 定义但附带条件 $\beta_{jr} = 0$,

$$Z_{ij} = \sum_{s=1}^{S} \gamma_{is} W_{js} \qquad [3.17]$$

除此之外,这个模型与多类别 logit 模型完全一样。通常忽略个人特征的原因是没有个人数据。因此,条件 logit 模型可以有效地看做个人的集合体——此后称做“总体”——在具有不同属性的选项之间做选择——比如,上下班的人选择不同的交通方式或者消费者选择不同的超级市场。通过这种想法,与个人相关的下标 i 就可以在下面的讲解中去掉了。因变量 $Y = j$ 与总体所做的选择相关,系数 γ_s 指总体附于第 s 个属性的权重。与在多类别模型中一样,在条件 logit 模型中任意两个选项 j 和 k 的风险比独立于其他可供选择的选项,

$$\log\left(\frac{\Pr(Y_i = j)}{\Pr(Y_i = k)}\right) = Z_j - Z_k = \sum_{s=1}^{S} (W_{js} - W_{ks})\gamma_s \qquad [3.18]$$

在方程 3.18 中,风险比对数仅仅取决于与 j 和 k 相关的

属性(间的差别),与其他选项的属性无关。[60]这个特性称为不相关选项的独立性(IIA),既是这些模型的主要优点,也是它们的主要缺点。[61]它是优点因为它允许在不需要重新估计模型的情况下引进新的选项。

比如,假设引进了第$(M+1)$个选项,其属性为 $W_{M+1,s}s = 1, \cdots, S$。与这个新选项相关的估计到的 Z 值(Z-value)为:$\hat{Z}_{M+1} = \sum_{s=1}^{S} \hat{\gamma}_s W_{M+1,s}$,其中系数估计值 $\hat{\gamma}_s$ 已经有了。新选项的采用所需要的只是,在方程 3.3 的分子中加入一个新的项 $\hat{Z}_{M+1}$并对概率重新计算。已有的系数估计值能同时在选项扩充之后使用的原因在于,新选项的加入并不能改变已有的选项被选择时的相对风险。

假定加入了某个新选项,选择特定选项的概率其变化的百分比可以很容易计算为:

$$\frac{\Pr(Y^{M+1}=j)-\Pr(Y^{M}=j)}{\Pr(Y^{M}=j)} = \left(\frac{\exp(Z_j)}{\sum_{j=1}^{M+1}\exp(Z_j)} \frac{\sum_{j=1}^{M}\exp(Z_j)}{\exp(Z_j)}\right)-1$$

$$= \frac{-\exp(Z_{M+1})}{\sum_{j=1}^{M+1}\exp(Z_j)}$$

$$= -\Pr(Y^{M+1}=M+1)$$

其中,$\Pr(Y^{M+1}=j)$ 和 $\Pr(Y^{M}=j)$ 分别指当有 $M+1$ 个和 M 个可供选择的选项时,选择选项 j 的概率。因此引进一个额外的选项时,选择已有的不同选项的概率将会减少相同的百分比,这个百分比等于选择这个新选项的概率。

但是,这个在引进新选项时已有概率下降相同百分比的特性同时也是一个缺点,因为它意味着对已有选项,关于新

选项的需求的交叉弹性在不同的选项间是一致的。[62]要使这个有效,这些选项必须被看做完全不同且独立的。它们被如此看待并导致IIA这个实际情况,源于误差是独立同分布这个假设。

该模型的局限性有多大的负面影响取决于所分析问题的本质。比如,在职业选择的模型中,选项可能被视为定义明确且相对不可改变,引进新选项带来的问题可能并不严重。[63]但另一方面,在选择上下班交通方式这个例子中,必须考虑到新的交通方式的引进——新的火车路线、公共汽车路线、小车交通拥挤费,等等——极有可能带来严重的问题。

一个典型的由IIA假设引起的问题的例子是"红巴士—蓝巴士"问题。假设上下班者有三种选择:乘坐小汽车、火车或者巴士去上班。顺便地假设所有的巴士都漆成红色,并且$\Pr(Y=j)$的logit估计值为:

小汽车:55.4

巴士(红色):23.0

火车:21.6

小汽车和巴士这两个交通方式间的风险比为55.4/23.0 = 2.4。现在巴士公司决定把一半的车漆成蓝色,我们可以很合理地预期这个纯粹表面的改变并不会给通勤者的选择带来任何变化,并且相同百分比的人,如上所示,将会继续使用小汽车、巴士和火车上班。但是,由于模型不能识别虚假的和真实的新选项,在引进"第四种"选项后预测到的logit概率变成[64]:

小汽车:45.1

巴士(红色):18.8

巴士(蓝色):18.4

火车:17.7

注意,每个已有选项小汽车、(红色)巴士和火车的概率都下降了18.4%,即使用蓝巴士上班的概率。使用小汽车和红色巴士(以及使用火车和红色巴士)之间的风险比保持不变,但是为了迁就这个情况,一些之前使用小汽车、火车或者红色巴士的通勤者必须改变以乘坐蓝色巴士上班。因此,仅仅由于把一半的车改漆成蓝色,巴士公司就可以使所有通勤者中乘坐巴士上班的比例由23%增加到37.2%。

另一个更明显的例子可以从政界中找到。假设某个政党 X 与其他两个政党 Y 和 Z 竞争。用多类别 logit 模型我们可以用相对于投 X 政党票的概率的一个比值,来估计支持 Z 的投票者的概率。假设现在政党 X 分成两个党,$X1$ 和 $X2$,它们具有类似的(比如左派的)观点,且笼统地被投票者认为两者间并没有本质差别。但是,由于模型不能识别虚假的和真实的新选项,在引进"第四种"政党后,预测到的投左派政党的票的 logit 概率将会大幅上升。这就是条件 logit 的一大缺陷:在 IIA 假设下,模型无法解决由引进类似的或相同的选项所带来的虚假的比例膨胀问题。

顺便提一下,我们应该注意由混合选择所带来的更进一步的局限性——比如,对上班时间和交通模式两者同时进行的选择。那么,正如多门西克和麦克法登(Domencich & McFadden, 1996)指出的,选择时间 t 和模式 j 的联合概率 $\Pr(t \cap j)$,可以写成(在时间 t,从可供利用的选择模式中,选择模式 j 的概率)和(从可供利用的时间选项中选择时间 t 的概率)的联合结果。换句话说,IIA 的一个含义是,包含混合选

择的效用函数必须是可以从个人选择中加法可分的，$U(t, j) = \phi(j) + \psi(t)$。

条件 logit 模型的替代模型

在考虑条件 logit 模型的替代方法之前，重要的是先检查 IIA 假设是否有效。假设我们相信选择的某个子集是不相关的。把它从选择集中去掉应该不会显著改变系数估计值。这个想法是豪斯曼和麦克法登（Hausman & McFadden, 1984）检验的依据。如果 $\hat{\gamma}_1$ 是基于受了限制的选择项而得到的系数估计值的向量，而 $\hat{\gamma}_0$ 是基于完整选项而得出的系数估计值的向量，且 $\hat{V}_1 - \hat{V}_0$ 是它们协方差矩阵的各自的估计值，那么统计值 $(\hat{\gamma}_1 - \hat{\gamma}_0)'(\hat{V}_1 - \hat{V}_0)(\hat{\gamma}_1 - \hat{\gamma}_0)$ 在 IIA 的限制是有效的这个零假设下，服从 $\chi^2(S)$分布。

如果这个零假设不被接受，那么一种可选择的方法是把方程 3.1 中的随机效用函数确定为多变量 probit 模型，$U_j = \sum_{s=1}^{S} \gamma_s W_{js} + \varepsilon_j \quad \varepsilon_j \sim N(0, \Sigma)$。其中，因为具有一个非纯量协方差矩阵 Σ，误差项就不再需要是独立的了。这个模型的问题是对多重正态积分和非限制协方差矩阵的估计进行计算这个实践问题(Greene, 2000:865)。

另一个放宽 IIA 限制的方法是对可供选择的选项进行分组，以使方差在组间不同，但在组内相同。这样就可以在选项组间放宽 IIA 假设，但是在每一组内仍然保留。这就是嵌套 logit 方法。这个问题不在这里讨论，详情及其应用请参考格林(1995)的论述。

第4章

STATA 程序列表

第 1 节 | 简介

本章列出了产生前文所讨论的结果的 STATA 程序清单。在概论部分作者已经讲过——但值得重复——至少有四种众所周知的具高度好评的软件,除了其他方面,还解决与本书所讨论的类似的问题:SAS, SPSS v10.0(印第安纳大学统计与数学计算中心为在次序与非次序结果事件的分析中应用 SAS 和 SPSS 提供了很好的入门课程[65]),LIMDEP (Greene, 1995),以及 STATA(STATA, 1999)。碰巧作者对 STATA 最熟悉,因此,这一章中的程序均用 STATA 编写。

下一部分包含了在第 2 章次序 logit 和 probit 模型中用到的程序清单,随后也包含了在第 3 章多类别 logit 模型中用到的程序清单。这些程序与相应的注释一起列出。这些注释的目的有三个:

(1) 解释特定的程序编码用来做什么(或做了什么);

(2) 把该编码与前面内容涉及的某个特定的表格联系起来,以使读者看到那些结果是如何产生的;

(3) 把该编码与前面内容涉及的某些特定的方程联系起来,以使读者看到用以产生某个特定结果的方法。

第2节 | 次序 probit 和 logit 程序

version 6.0 / * 使用 STATA 版本 6.0 * /;

use C:\SAGE\NLdta / * 读取 STATA 格式的数据:58 个变量,13164 个观察值 * /;

#delimit; / * 命令将由;终止 * /;

/ * 题目:使用剥夺分析例子的次序 logit 和 probit * /;

gen pnum = _N; / * 对每个人分配一个数字 * /;

/ * 对次序 logit, y 是因变量;y = 1 为没有被剥夺;y = 2 为轻度被剥夺; y = 3 为严重被剥夺 * /;

tabulate y rc, col; tabulate y sex, col;

/ * 用整个次样本估计 Oprobit 方程:表 2.2 * /;

oprobit y sex ct age age2 ret inac ue highed mided hnum snpar ard dwn crk ant col arm ban dry frm, table;

predict p1 p2 p3; / * 估计到的概率储存为 p1, p2, p3 * /;

sort pnum; / * 观察值用个人的数字以升序的形式分类 * /;

list pnum p1 p2 p3 in 1/25, noobs; / * 用个人数字列出前 25 个人的估计到的概率:表 2.4 * /;

summarize p1 p2 p3, detail; / * 详细概括估计到的个人

概率:表 2.6 */;

predict z, xb; /*对每个人计算 Z 值*/;

summarize z, mean; gen zm = r(mean); /*计算 Z 的均值并存于 zm 这个变量中*/;

/* b[_cut1] and b[_cut2]储存了临界值 δ_1 和 δ_2 的估计值*/;

gen p1 = normprob(_b[_cut1] − z); gen p2 = normprob(_b[_cut2] − z) − normprob(_b[_cut1] − z); gen p3 = 1 − normprob(_b[_cut2] − z);

/* 对每个人 p1, p2 和 p3 为用方程 2.5a 至 2.5c 计算出来的概率的估计值,这些值与之前计算出来的 p1 − p3 的估计值相同,上面*/;

gen q1 = normprob(_b[_cut1] − zm); gen q2 = normprob(_b[_cut2] − zm) − normprob(_b[_cut1] − zm); gen q3 = 1 − normprob(_b[_cut2] − zm);

/*q1, q2 和 q3 是在设定取值等于样本均值的情况下计算出来的概率估计值:方程 2.18a 至 2.18c */;

summarize p1 p2 p3; summarize q1 q2 q3; /*比较概率:表 2.8 */;

drop p1 p2 p3 q1 q2 q3; /*因为随后还要用到,在此放弃这些变量名*/

/*计算边际效应:连续变量(Age)*/;

gen a1 = normd(_b[_cut1] − z) * _b[age];

gen a2 = (normd(_b[_cut2] − z) − normd(_b[_cut1] − z) * _b[age];

gen a3 =− 1 * normd(_b[_cut2] − z) * _b[age];

/ * a1 − a3 是对每个人计算的 AGEi 的边际效应:方程 2. 13a 至 2. 13c * /;

gen am1 = normd(_b[_cut1] − zm) * _b[age]; gen am2 = (normd(_b [_cut2] − zm) − normd(_b[_cut1] − zm)) * _b[age];

gen am3 =− 1 * normd(_b[_cut2] − zm) * _b[age];

/ * am1 − am3 是设定变量值等于样本均值时 AGE 的边际效应 * /;

gen b1 = normd(_b[_cut1] − z) * _b[age2];

gen b2 = (normd(_b[_cut2] − z) − normd(_b[_cut1] − z)) * _b[age2];

gen b3 =− 1 * normd(_b[_cut2] − z) * _b[age2];

/ * b1 − b3①是对每个人计算的 AGE_i^2 的边际效应:方程 2. 13a 至2. 13c * /;

gen bm1 = normd(_b[_cut1] − zm) * _b[age2];

gen bm2 = (normd(_b[_cut2] − zm) − normd(_b[_cut1] − zm)) * _b[age2];

gen bm3 =− 1 * normd(_b[_cut2] − zm) * _b[age2];

/ * bm1 − bm3②是设定变量值等于样本均值时 AGE^2 的边际效应 * /;

gen c1 = a1 + b1; gen c2 = a2 + b2; gen c3 = a3 + b3; gen cm1 = am1 + bm1; gen cm2 = am2 + bm2; gen cm3 = am3 + bm3;

/ * 加上 AGE 和 AGE^2 的作用 * /;

---

① 原文中为 a1 − a3,其实有误。——译者注

② 原文中为 am1 − am3,其实有误。——译者注

summarize c1 c2 c3; summarize cm1 cm2 cm3; /* 表 2.10 */;

drop z zm c1 a1 b1 c2 a2 b2 c3 a3 b3 cm1 am1 bm1 cm2 am2 bm2 cm3 am3 bm3;

/* 计算边际效应:虚拟变量(Religion) */;

gen cto = ct; /* 保存原值 */;

replace ct = 1; /* 每个人都是天主教徒 */;

predict p1 p2 p3; /* 当 CTi = 1 对每个人估计概率 */;

predict z, xb; summarize z; gen zm = r(mean); /* 计算 Z 的均值并存于 zm 这个变量中 */;

gen q1 = normprob(_b[_cut1] − zm); gen q2 = normprob(_b[_cut2] − zm) − normprob(_b[_cut1] − zm); gen q3 = 1 − normprob(_b[_cut2] − zm);

/* q1, q2 和 q3 是在设定非宗教取值等于样本均值的情况下计算出来的概率估计值 */;

summarize p1 p2 p3; summarize q1 q2 q3; /* 表 2.12 */;

drop p1 p2 p3 q1 q2 q3 z zm;

replace ct = 0; /* 每个人都是新教徒 */;

predict p1 p2 p3; /* 当 CTi = 0 对每个人估计概率 */;

predict z, xb; summarize z; gen zm = r(mean); /* 计算 Z 的均值并存于 zm 这个变量中 */;

gen q1 = normprob(_b[_cut1] − zm); gen q2 = normprob(_b[_cut2] − zm) − normprob(_b[_cut1] − zm); gen q3 = 1 − normprob(_b[_cut2] − zm);

/＊ q1，q2 和 q3 是在设定非宗教取值等于样本均值的情况下计算出来的概率估计值＊/；

summarize p1 p2 p3； summarize q1 q2 q3； /＊ 表 2.12 ＊/；

drop p1 p2 p3 q1 q2 q3 z zm; replace ct = cto; /＊恢复原值＊/；

/＊用整个次样本估计 Oprobit 方程：表 2.1 ＊/；

ologit y sex ct age age2 ret inac ue highed mided hnum snpar ard dwn crk ant col arm ban dry frm，table;

predict p1 p2 p3； /＊ 估计到的概率储存为 p1，p2，p3＊/；

sort pnum; /＊观察值用个人的数字以升序的形式分类＊/；

list pnum p1 p2 p3 in 1/25，noobs; /＊用个人数字列出前 25 个人的估计到的概率：表 2.4＊/；

summarize p1 p2 p3，detail; /＊详细概括估计到的个人概率：表 2.6＊/；

predict z，xb; /＊对每个人计算 Z 值＊/；

summarize z，mean; gen zm = r(mean)；/＊计算 Z 的均值并存于 zm 这个变量中＊/；

/＊ b[_cut1] 和 b[_cut2] 储存了临界值 δ_1 和 δ_2 的估计值＊/；

gen p1 = 1/(1＋exp(z－_b[_cut1]))；gen p2 = 1/(1＋exp(z－_b[_cut2]))－1/(1＋exp(z－_b[_cut1]))；gen p3 = 1－1/(1＋exp(z－_b[_cut2]))；

/＊对每个人 p1，p2 和 p3 为用方程 2.5a 至 2.5c 计算

出来的概率的估计值,这些值与之前计算出来的 p1 − p3 的估计值相同,上面 * /;

gen q1 = 1/(1+exp(zm−_b[_cut1])); gen q2 = 1/(1+exp(zm−_b[_cut2])) − 1/(1+exp(zm−_b[_cut1])); gen q3 = 1 − 1/(1+exp(zm−_b[_cut2]));

/* q1, q2 和 q3 是在设定取值等于样本均值的情况下计算出来的概率估计值:方程 2.18a 至 2.18c * /;.

summarize p1 p2 p3; summarize q1 q2 q3; / * 比较概率:表 2.7 * /;

drop p1 p2 p3 q1 q2 q3; / * 因为随后还要用到,在此放弃这些变量名 * /;

/ * 计算边际效应:连续变量(Age) * /;

gen lden1 = 1/(1+exp(z−_b[_cut1]));

gen lden2 = 1/(1+exp(z−_b[_cut2]));

gen lden3 = 1/(1+exp(z−_b[_cut2]));

/ * logit 概率计算出来了:方程 2.9a 至 2.9c * /;

gen a1 = lden1 * (1−lden1) * _b[age]; gen b1 = lden1 * (1 − lden1) * _b[age2]; gen c1 = a1 + b1; gen a2 = (lden2 * (1 − lden2) − lden1 *

(1 − lden1)) * _b[age];

gen b2 = (lden2 * (1 − lden2) − lden1 * (1 − lden1)) * _b[age2];

gen c2 = a2+b2;

gen a3 =− lden3 * (1 − lden3) * _b[age];

gen b3 =− lden3 * (1 − lden3) * _b[age2];

gen c3 = a3 + b3;

/* a1 — a3 是对每个人计算的 AGEi 的边际效应:方程 2. 12a 至 2. 12c */;

/* b1 — b3 是对每个人计算的 AGE_i^2 的边际效应:方程 2. 12a 至2. 12c */;

/* c1 — c3 是这些效应的和 */;

replace lden1 = 1/(1 + exp(zm — _b[_cut1]));

replace lden2 = 1/(1 + exp(zm — _b[_cut2]));

replace lden3 = 1/(1 + exp(zm — _b[_cut2]));

/* logit 概率用均值计算 */;

gen am1 = lden1 * (1 — lden1) * _b[age];

gen bm1 = lden1 * (1 — lden1) * _b[age2];

gen cm1 = am1 + bm1;

gen am2 = (lden2 * (1 — lden2) — lden1 * (1 — lden1)) * _b[age];

gen bm2 = (lden2 * (1 — lden2) — lden1 * (1 — lden1)) * _b[age2];

gen cm2 = am2 + bm2;

gen am3 =— lden3 * (1 — lden3) * _b[age];

gen bm3 =— lden3 * (1 — lden3) * _b[age2];

gen cm3 = am3 + bm3;

/* am1 — am3[①]是用样本均值计算的 AGEi 的边际效应 */; /* bm1 — bm3[②]是用样本均值计算的 AGE 的边际效应 */; /* cm1 — cm3[③]是这些效应的和 */;

---

① 原为 a1 — a3,其实有误。——译者注

② 原为 b1 — b3,其实有误。——译者注

③ 原为 c1 — c3,其实有误。——译者注

summarize c1 c2 c3; summarize cm1 cm2 cm3; /* 表2.9 */;

drop z zm c1 a1 b1 c2 a2 b2 c3 a3 b3 cm1 am1 bm1 cm2 am2 bm2 cm3 am3 bm3;

replace ct = 1; /* 每个人是天主教徒 */;

predict p1 p2 p3; /* 当CTi = 1对每个人估计概率 */; predict z, xb; summarize z; gen zm = r(mean); /* 计算Z的均值并存于zm这个变量中 */;

gen q1 = 1/(1+exp(zm-_b[_cut1])); gen q2 = 1/(1+exp(zm-_b[_cut2]))-1/(1+exp(zm-_b[_cut1])); gen q3 = 1-1/(1+exp(zm-_b[_cut2]));

/* q1, q2和q3是在设定非宗教取值等于样本均值的情况下计算出来的概率估计值 */;

summarize p1 p2 p3; summarize q1 q2 q3; /* 表2.11 */;

drop p1 p2 p3 q1 q2 q3 z zm;

replace ct = 0; /* 每个人都是新教徒 */;

predict p1 p2 p3; /* 当CTi = 0对每个人估计概率 */;

predict z xb; summarize z; gen zm = r(mean); /* 计算Z的均值并存于zm这个变量中 */;

gen q1 = 1/(1+exp(zm-_b[_cut1])); gen q2 = 1/(1+exp(zm-_b[_cut2]))-1/(1+exp(zm-_b[_cut1])); gen q3 = 1-1/(1 +exp(zm-_b[_cut2]));

/* q1, q2和q3是在设定非宗教取值等于样本均值的情况下计算出来的概率估计值 */;

summarize p1 p2 p3; summarize q1 q2 q3; /* 表

2.11 */;

drop p1 p2 p3 q1 q2 q3 z zm;

replace ct = cto; /*恢复原值*/;

/*用新教徒次样本估计Oprobit方程*/;

ologit y sex age age2 ret inac ue highed mided hnum snpar ard dwn crk ant col arm ban dry frm if ct = 0, table;

lrtest,saving(0); /*为LR检验储存概率值*/;

ologit y sex age age2 ret inac ue highed mided hnum snpar ard crk ant col arm ban frm if ct = 0, table;

lrtest; /*LR检验dwn和dry的系数联合等于0*/;

/*对新教徒用新教徒的系数进行预测,且计算均值*/;

predict ppa1 ppa2 ppa3 if ct = 0; egen mppa1 = mean(ppa1); egen mppa2 = mean(ppa2); egen mppa3 = mean(ppa3);

predict pps1 pps2 pps3 if ct = 0 & snpar = 1;

egen mpps1 = mean(pps1); egen mpps2 = mean(pps2); egen mpps3 = mean(pps3); predict ppr1 ppr2 ppr3 if ct = 0 & ret = 1; egen mppr1 = mean(ppr1); egen mppr2 = mean(ppr2); egen mppr3 = mean(ppr3);

predict ppi1 ppi2 ppi3 if ct = 0 & inac = 1;

egen mppi1 = mean(ppi1); egen mppi2 = mean(ppi2); egen mppi3 = mean(ppi3); predict ppu1 ppu2 ppu3 if ct = 0 & ue = 1; egen mppu1 = mean(ppu1); egen mppu2 = mean(ppu2); egen

mppu3 = mean(ppu3);

predict ppl1 ppl2 ppl3 if ct = 0 & resnum = 1;

egen mppl1 = mean(ppl1); egen mppl2 = mean(ppl2); egen mppl3 = mean(ppl3);

/ * 用天主教徒次样本估计 Oprobit 方程 * /;

ologit y sex age age2 ret inac ue highed mided hnum snpar ard dwn crk ant col arm ban dry frm if ct = 1, table;

lrtest, saving(0); / * 为 LR 检验储存概率值 * /;

ologit y sex age age2 ret inac ue highed mided hnum snpar ard crk ant col arm ban frm if ct = 1, table; lrtest; / * LR 检验 dwn 和 dry 的系数联合等于 0 * /;

/ * 对天主教徒用天主教徒的系数进行预测,且计算均值 * /;

predict cca1 cca2 cca3 if ct = 1;

egen mcca1 = mean(cca1); egen mcca2 = mean(cca2); egen mcca3 = mean(cca3); predict ccs1 ccs2 ccs3 if ct = 1 & snpar = 1; egen mccs1 = mean(ccs1); egen mccs2 = mean (ccs2); egen mccs3 = mean(ccs3);

predict ccr1 ccr2 ccr3 if ct = 1 & ret = 1;

egen mccr1 = mean(ccr1); egen mccr2 = mean(ccr2); egen mccr3 = mean(ccr3); predict cci1 cci2 cci3 if ct = 1 & inac = 1; egen mcci1 = mean(cci1); egen mcci2 = mean (cci2); egen mcci3 = mean(cci3);

predict ccu1 ccu2 ccu3 if ct = 1 & ue = 1;

egen mccu1 = mean(ccu1); egen mccu2 = mean(ccu2); egen mccu3 = mean(ccu3);

predict cel1 cel2 cel3 if ct = 1 & resnum = 1;

egen mcel1 = mean(ccl1); egen mcel2 = mean(cel2); egen mcel3 = mean(cel3);

/*对新教徒用天主教徒的系数进行预测,且计算均值*/;

predict pca1 pca2 pca3 if ct = 0;

egen mpca1 = mean(pca1); egen mpca2 = mean(pca2); egen mpca3 = mean(pca3);

predict pcs1 pcs2 pcs3 if ct = 0 & snpar = 1;

egen mpcs1 = mean(pcs1); egen mpcs2 = mean(pcs2); egen mpcs3 = mean(pcs3);

predict pcr1 pcr2 pcr3 if ct = 0 & ret = 1;

egen mpcr1 = mean(pcr1); egen mpcr2 = mean(pcr2); egen mpcr3 = mean(pcr3);

predict pci1 pci2 pci3 if ct = 0 & inac = 1;

egen mpci1 = mean(pci1); egen mpci2 = mean(pci2); egen mpci3 = mean(pci3);

predict pcu1 pcu2 pcu3 if ct = 0 & ue = 1;

egen mpcu1 = mean(pcu1); egen mpcu2 = mean(pcu2); egen mpcu3 = mean(pcu3);

predict pcl1 pcl2 pcl3 if ct = 0 & resnum = 1;

egen mpcl1 = mean(pcl1); egen mpcl2 = mean(pcl2); egen mpcl3 = mean(pcl3);

/*现在百分比贡献(percentage contributions)可以计算出来了*/;

gen Aa1 = [(mppa1 − mpca1)/(mppa1 − mcca1)] *100;

```
gen Aa2 = [(mppa2 - mpca2)/(mppa2 - mcca2)]
* 100;
gen Aa3 = [(mppa3 - mpca3)/(mppa3 - mcca3)]
* 100;
gen As1 = [(mpps1 - mpcs1)/(mpps1 - mccs1)]
* 100;
gen As2 = [(mpps2 - mpcs2)/(mpps2 - mccs2)]
* 100;
gen As3 = [(mpps3 - mpcs3)/(mpps3 - mccs3)]
* 100;
gen Ar1 = [(mppr1 - mpcr1)/(mppr1 - mccr1)] * 100
gen Ar2 = [(mppr2 - mpcr2)/(mppr2 - mccr2)]
* 100;
gen Ar3 = [(mppr3 - mpcr3)/(mppr3 - mccr3)]
* 100;
gen Ai1 = [(mppi1 - mpci1)/(mppi1 - mcei1)] * 100;
gen Ai2 = [(mppi2 - mpci2)/(mppi2 - mcci2)] * 100;
gen Ai3 = [(mppi3 - mpci3)/(mppi3 - mcci3)] * 100;
gen Au1 = [(mppu1 - mpcu1)/(mppu1 - mccu1)]
* 100;
gen Au2 = [(mppu2 - mpcu2)/(mppu2 - mccu2)]
* 100;
gen Au3 = [(mppu3 - mpcu3)/(mppu3 - mccu3)]
* 100;
gen Al1 = [(mppl1 - mpel1)/(mppl1 - mcel1)] * 100;
gen Al2 = [(mppl2 - mpel2)/(mppl2 - mcel2)] * 100;
```

```
gen Al3 = [(mppl3 - mpel3)/(mppl3 - mcel3)] * 100;
summarize
ppa1 ppa2 ppa3 pca1 pca2 pca3 cca1 cca2 cca3 pps1 pps2
pps3 pcs1 pcs2 pcs3 ccs1 ccs2 ccs3 ppr1 ppr2 ppr3 pcr1 pcr2
pcr3 ccr1 ccr2 ccr3 ppi1 ppi2 ppi3 pci1 pci2 pci3 cci1 cci2 cci3
ppu1 ppu2 ppu3 pcu1 pcu2 pcu3 ccu1 ccu2 ccu3 ppl1 ppl2 ppl3
pcl1 pcl2 pcl3 ccl1 ccl2 ccl3;
/*表2.16 */;
summarize Aa1 Aa2 Aa3 As1 As2 As3 Ar1 Ar2 Ar3 Ai1
Ai2 Ai3 Au1 Au2 Au3 Al1 Al2 Al3; /*表2.17 */;
/*程序结束*/;
```

第 3 节 | 多类别 logit 程序

use c:\SAGE\GB. dta / * 读取 STATA 格式的数据:24 个变量,98732 个观察值:这些数据为白人、黑人和印第安人全职男性雇员,25—45 岁 * /;

delimit;

/ * 题目:使用职业类别分析例子的多类别 logit * /;

gen pnum = _N; / * 对每个人分配一个数字 * /;

/ * 多类别 logit 中 y 是因变量,且

y = 1 如果这个人受雇于非技术/半技术工人职业类别

y = 2 如果这个人受雇于有技术体力/非体力劳动者职业类别

y = 3 如果这个人受雇于专业/管理/技术人员职业类别 * /;

/ * 对所有人、白人、黑人、印第安人列表 * /;

tab y; tab y if blk = 0 &ind = 0; tab y if blk = 1; tab y if ind = 1;

/ * 现在估计多类别 logit 方程,表 3. 2:模型设定见方程 3. 14。注:以下 base(1)设结果 1 为基准值 * /;

mlogit y age age2 blk blkovs ind indovs ovsbn north indnth

blknth south indsth blksth highed indhe blkhe mided

indme blkme subbus subsei bush scih, base(1);

lrtest, saving(0); /* 为 LR 检验储存似然对数 */;

/* 确定零限制:对 y = 2,blk 和 blkovs 从方程中去掉;对 y = 3,ind, blknth, scih, bush, subsci 从方程中去掉;indme, indhe, blkme 在两个方程中都去掉 */;

constraint define 1 [2]blkovs = 0; constraint define 2 [3]blk = 0;

constraint define 3 [3]ind = 0; constraint define 4 [3] blknth = 0;

constraint define 5 [3]scih = 0; constraint define 6 [3] bush = 0;

constraint define 7 [3]subsci = 0; constraint define 8 indhe;

constraint define 9 indme; constraint define 10 blkme;

/* 现在在强加限制 1—10 的情况下估计 mlogit 方程,表 3.3。注:下面的 constr(1—10)为强加限制的命令 */;

mlogit y age age2 blk blkovsind indovs ovsbn north indnth

blknth south indsth blksth highed indhe blkhe mided indme

blkme subbus subsci bush scih, constr(1—10) base(1);

lrtest; /* 似然比检验完成:chi2(13) = 8.1 零限制没有被拒绝。 */;

/* 对每个人对结果 1, 2, 3 估计到的概率分别储存为 p1, p2, p3 */;

predict p1, outcome(1); predict p2, outcome(2); pre-

dict p3, outcome(3);

/* 对所有人,白人,黑人和印第安人,概括概率估计值(表 3.4 上部分) */;

summarize p1 p2 p3; summarize p1 p2 p3 if blk. = 0 & ind = 0; summarize p1 p2 p3 if blk = 1; summarize p1 p2 p3 if ind = 1;

drop p1 p2 p3; /* 因为随后还要用到,在此放弃这些变量名 */;

/* 对每个人对结果 1, 2, 3 计算 Z 值,并储存为 z1, z2, z3 */;

predict z1, outcome(1) xb; predict z2, outcome(2) xb; predict z3, outcome(3) xb;

/* z1, z2, z3 的均值储存为 zm1, zm2, zm3 */;

summarize z1, mean; gen zm1 = r(mean); summarize z2, mean;

gen zm2 = r(mean); summarize z3, mean; gen zm3 = r (mean);

/* 对所有人,用方程 3.12a 和 3.12b 计算结果 1, 2, 3 的概率 */;

gen sum = 1+exp(zm2)+exp(zm3); gen p1 = 1/sum;

gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum; summarize p1 p2 p3; /* 对所有人展示了概率估计值(表 3.4 下部分) */;

drop sum p1 p2 p3 zm1 zm2 zm3; /* 因为随后还要用到,在此放弃这些变量名 */;

/* 现在对白人计算 z1 z2 z3 的均值,并储存于 zm1,

zm2, zm3 */;

summarize z1 if blk = 0 & ind = 0, mean; gen zm1 = r(mean); summarize z2 if blk = 0 & ind = 0, mean; gen zm2 = r(mean); summarize z3 if blk = 0 & ind = 0, mean; gen zm3 = r(mean);

/*对白人,用方程 3.12a 和方程 3.12b 计算结果 1, 2, 3 的概率 */;

gen sum = 1+exp(zm2)+exp(zm3); gen p1 = 1/sum; gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum; summarize p1 p2 p3 if blk = 0 &ind = 0;

/*对白人展示了概率估计值(表 3.4 下部分)*/;

drop sum p1 p2 p3 zm1 zm2 zm3; /*因为随后还要用到,在此放弃这些变量名 */;

/*现在对黑人计算 z1 z2 z3 的均值,并储存于 zm1, zm2, zm3 */;

summarize z1 if blk = 1, mean; gen zm1 = r(mean);

summarize z2 if blk = 1, mean; gen zm2 = r(mean);

summarize z3 if blk = 1, mean; gen zm3 = r(mean);

/*对黑人,用方程 3.12a 和方程 3.12b 计算结果 1, 2, 3 的概率 */;

gen sum = 1+exp(zm2)+exp(zm3); gen p1 = 1/sum; gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum;

summarize p1 p2 p3 if blk = 1; /*对黑人展示了概率估计值(表 3.4 下部分)*/;

drop sum p1 p2 p3 zm1 zm2 zm3; /* Variables are released for subsequent use */;

/＊现在对印第安人计算 z1 z2 z3 的均值,并储存于 zm1, zm2, zm3＊ /;

summarize z1 if ind = 1, mean; gen zm1 = r(mean);

summarize z2 if ind = 1, mean; gen zm2 = r(mean);

summarize z3 ifind = 1, mean; gen zm3 = r(mean);

/＊对印第安人,用方程 3.12a 和方程 3.12b 计算结果 1, 2, 3 的概率＊/;

gen sum = 1 + exp(zm2) + exp(zm3); gen p1 = 1/sum; gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum;

summarize p1 p2 p3 if ind = 1; /＊对印第安人展示了概率估计值(表 3.4 下部分)＊/;

drop sum p1 p2 p3 zm1 zm2 zm3; drop z1 z2 z3; /＊因为随后还要用到,在此放弃这些变量名＊ /;

/＊接着为种族模拟分析＊ /;

gen blko = blk; gen indo = ind; /＊保存种族变量＊/;

replace blk = 0; replace ind = 0; /＊每个人都是白人＊ /;

/＊所有涉及 blk 和 ind 的交互变量都需要重新计算＊ /;

replace indnth = ind ＊ north; replace indsth = ind ＊ south;

replace blknth = blk ＊ north; replace blksth = blk ＊ south;

replace blkovs = blk ＊ ovsbn; replace indovs = ind ＊ ovsbn;

replace blkhe = blk ＊ highed;

/＊对每个人对结果 1, 2, 3 估计到的概率分别储存为 p1, p2, p3＊/;

predict p1, outcome(1); predict p2, outcome(2); predict p3, outcome(3); summarize p1 p2 p3;

/ * 对整个样本概括概率估计值:每个人都假定为白人(表 3.5 上部分) * /;

drop p1 p2 p3;

/ * 对每个人对结果 1, 2, 3 计算 Z 值并储存于 z1, z2, z3: 每个人都假定为白人 * /;

predict z1, outcome(1) xb; predict z2, outcome(2) xb; predict z3, outcome(3) xb;

/ * 对整个样本,计算 z1 z2 z3 的均值,并储存于 zm1, zm2, zm3 * /;

summarize z1, mean; gen zm1 = r(mean); summarize z2, mean; gen zm2 = r(mean); summarize z3, mean; gen zm3 = r(mean);

/ * 用方程 3.12a 和方程 3.12b 计算结果 1, 2, 3 的概率:每个人都假定为白人 * /;

gen sum = 1 + exp(zm2) + exp(zm3); gen p1 = 1Isum; gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum;

summarize p1 p2 p3; / * 对整个样本概括概率估计值:每个人都假定为白人(表 3.5 下部分) * /;

drop z1 z2 z3 sum p1 p2 p3 zm1 zm2 zm3; / * 因为随后还要用到,在此放弃这些变量名 * /;

replace blk = 1; / * 每个人都是黑人 * /;

/ * 所有涉及 blk 和 ind 的交互变量都需要重新计算 * /;

replace blknth = blk * north; replace blksth = blk * south;

replace blkovs = blk * ovsbn; replace blkhe = blk * highed;

/ * 没有必要替换 ind * 变量因为 ind 已经等于 0 * /;

/ * 对每个人对结果 1, 2, 3 估计到的概率分别储存为 p1, p2, p3 * /;

Predict p1, outcome(1); predict p2, outcome(2); predict p3, outcome(3);

summarize p1 p2 p3; drop p1 p2 p3; / * 对整个样本概括概率估计值:每个人都假定为黑人(表 3.5 上部分) * /;

/ * 对每个人对结果 1, 2, 3 计算 Z 值并储存于 z1, z2, z3: 每个人都假定为黑人 * /;

predict z1, outcome(1) xb; predict z2, outcome(2) xb; predict z3, outcome(3) xb;

/ * 对整个样本,计算 z1 z2 z3 的均值,并储存于 zm1, zm2, zm3 * /;

summarize z1, mean; gen zm1 = r(mean); summarize z2, mean; gen zm2 = r(mean); summarize z3, mean; gen zm3 = r(mean);

/ * 用方程 3.12a 和方程 3.12b 计算结果 1, 2, 3 的概率:每个人都假定为黑人 * /;

gen sum = 1+exp(zm2)+exp(zm3); gen p1 = 1/sum;

gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum;

summarize p1 p2 p3; / * 对整个样本概括概率估计值:每个人都假定为黑人(表 3.5 下部分) * /;

drop z1 z2 z3 sum p1 p2 p3 zm1 zm2 zm3; / * 因为随后还要用到,在此放弃这些变量名 * /;

replace ind = 1; replace blk = 0; /* 每个人都是印第安人 */;

/* 所有涉及 blk 和 ind 的交互变量都需要重新计算 */;

replace indnth = ind * north; replace indsth = ind * south;

replace blknth = blk * north; replace blksth = blk * south;

replace blkovs=blk * ovsbn; replace indovs=ind * ovsbn;

replace blkhe=blk * highed;

/* 对每个人对结果 1, 2, 3 估计到的概率分别储存为 p1, p2, p3 */;

predict p1, outcome(1); predict p2, outcome(2); predict p3, outcome(3);

summarize p1 p2 p3; drop p1 p2 p3; /* 对整个样本概括概率估计值:每个人都假定为印第安人(表 3.5 上部分) */;

/* 对每个人对结果 1, 2, 3 计算 Z 值并储存于 z1, z2, z3: 每个人都假定为印第安人 */;

predict z1, outcome(1) xb; predict z2, outcome(2) xb; predict z3, outcome(3) xb;

/* 对整个样本,计算 z1 z2 z3 的均值,并储存于 zm1, zm2, zm3 */;

summarize z1, mean; gen zm1 = r(mean); summarize z2, mean; gen zm2 = r(mean); summarize z3, mean; gen zm3 = r(mean);

/＊用方程 3.12a 和方程 3.12b 计算结果 1，2，3 的概率:每个人都假定为印第安人＊/;

gen sum = 1 + exp(zm2) + exp(zm3); gen p1 = 1/sum;

gen p2 = exp(zm2)/sum; gen p3 = exp(zm3)/sum;

summarize p1 p2 p3; /＊对整个样本概括概率估计值:每个人都假定为印第安人表 3.5 下部分＊/;

drop z1 z2 z3 sum p1 p2 p3 zm1 zm2 zm3; /＊因为随后还要用到,在此放弃这些变量名＊/;

replace blk = blko; replace ind = indo; /＊恢复原来的变量＊/;

/＊程序结束＊/;

注释

[1] 在网站 http://www.indiana.edu/~statmath/stat/all/cat/giant.html 可以找到。

[2] 只有两种结果时，次序 logit 与 probit 这两种方法都可以使用，不管因变量是有序的还是非有序的。

[3] 但是，有序模型可以基于除 logit 或 probit 分布之外的其他分布。比如，对偏斜有序数据(skewed ordinal data)的双对数(log-log)、负双对数(negative log-log)或互补双对数(complementary log-log)分布(见 Agresti，1990)。大致要点是，次序 logit 和 probit 模型是一系列更广泛得多的处理有序因变量的观点中的一部分。其实，如果结果的数目很大(比如，大于 20)，那么次序 logit 和 probit 的方法就会变得很笨重，用其他方法可能会更合适(关于这点的讨论，见 Jöreskog & Sörbom，1988:44—45；1993:1—17)。

[4] 排序可以基于收入，这种情况下职业具有一个客观的等级；或者也可基于文化假设和偏见。

[5] 比"把结果当成是有序的，除非我们有好的理由不强加一个排序"更好。

[6] 记住，$N = N_1 + N_2 + N_3$。

[7] 因为它们的计算方法，这些估计值被称为最大似然估计值(maximum likelihood estimates)。

[8] 还有别的把有序结果模型化的方法，比如相邻类别(adjacent categories)logit 模型，构造型(stereotype)模型，logit 模型和连续比(continuation ratio)logit 模型(见 Agresti，1996:216—220)。

[9] 它非常类似于一个自由度为 7 的 t 分布(Greene，2000:815)。

[10] 关于这个属性的讨论见 Brant(1990)。

[11] 其实 Greene (2000)把它的临界值记为从 0 开始，所以在它的注释中，$\mu = 0$。但为了方便比较，我把他的临界值起点记为 1。

[12] 记住，Greene(2000:876)把 $\hat{W}_i$ 写成 $\beta' x$。

[13] 因为 $\mu_1 = 0$，$\delta_1 = -\beta_0$，所以 STATA 的第一个临界值是负的截距。

[14] 均值为 0 方差为 1 的正态分布。

[15] 记住，这三个概率之和必须为 1。

[16] 与其中一个结果相关的系数必须为了可识别性而被标准化。

[17] 这些人相当于这次人口普查记录的一个 2%样本。对这个数据的更完整的描述见 Borooah(2000a)。

[18] 某个人被认为没有被剥夺如果 $D_i = 0$，轻度被剥夺如果 $0 < D_i \leqslant \bar{D}$，

严重被剥夺如果 $D_i > \bar{D}$，其中 $\bar{D}$ 为剥夺指数的均值。

[19] 另一种标准化，设定17的人 $AGE_i = 1$，18岁的 $AGE_i = 2$，等等。

[20] 这些是通常在18岁或18岁以上获得的文凭。

[21] 他们是：Belfast（地区1）；Ards，Castlereagh，North Down（地区2）；Down，Lisburn（地区3）；Carrickfergus，Larne，Newtwoenabbey（地区4）；Antrim，Ballymena and Ballymoney（地区5）；Armagh，Newry & Mourne（地区7）；Coleraine，Cookstown，Maghrafelt，Moyle（地区6）；Banbridge，Craigavon，Dungannon（地区8）；Derry，Limavady（地区9）；and Fermanagh，Omagh，Strabane（地区10）。因为多重共线性，不能10地区全都放在方程2.16中。$AREA_1$（Belfast）这个地区从方程中去掉了。

[22] 所有非罗马天主教徒识别为"基督徒"，虽然后者还包括要么没有陈述宗教信仰（北爱尔兰居民的7.2%），要么宣称他们没有信仰的人（北爱尔兰居民的3.8%）。

[23] 注意，(a)在STATA里，$\beta_1 = 0$，(b)从方程2.16，$dD_i/dAGE_i = \beta_4 + 2\beta_5 AGE_i$。

[24] 这些比率表示沃尔德统计量（Wald statistic）的一种形式。

[25] 为了测量平行斜率假设，这个模型还用多类别logit进行了估计，产生了一个等于−12380.36的似然值（记为 L_2）。由 $2(L_2 - L_1)$ 计算出来的值为86.4（与次序logit模型进行的比较），依照严格的似然比检验的解释，这个值超过了 $\chi^2(20)$ 的5%临界值31.41。但是，正如前面指出的，$2(L_2 - L_1)$ 的值只是提示性的，因为这个统计值并没有提供似然比检验的基础。

[26] 说明：（表中的）$xb + u =$（文字中的）$\hat{Z}_i + \varepsilon_i$。

[27] 临界点其实只是模型的系数，且与斜率系数一起用最大似然方法估计而得。

[28] 这些比例在这个章节开头的时候就设定了。

[29] 虽然当平均数被定义为个人概率的中位数（而不是均值）时结果会不一样。

[30] 注意，在STATA中，$\beta_1 = 0$。

[31] 为了方便表述，随后的讨论全都针对logit模型。

[32] 比如，与8%的新教徒相比，样本中将近16%的天主教徒为单身父/母亲。

[33] 注意，这些预测到的均值会与天主教徒与新教徒处于不同剥夺程度的样本比例相同。

[34] 即新教徒的属性用新教徒的系数进行评估。

[35] 在失业者和退休者的情况下，这些概率几乎不变。

[36] 在具有相同的处于轻度被剥夺的平均概率的意义上来说。

[37] 这些人相当于这次人口普查记录的一个 2%样本。对这个数据的更完整的描述见 Borooah(2000b)。

[38] 缺省值为没有 18 岁或 18 岁以上的文凭。

[39] 只在这个人具有 18 岁或 18 岁以上的文凭时有意义。缺省值为文科类课程。

[40] 缺省值为白人。

[41] 缺省值为伦敦。

[42] 关于英国的标准区域划分：中英格兰东(East Midlands)；东英格利亚(East Anglia)；东南部(the Southeast)（伦敦除外）；西南部(the Southwest)。

[43] 北部；约克郡(Yorkshire)；中英格兰西(West Midlands)；西北部；威尔士(Wales)；苏格兰(Scotland)。

[44] 注意，这三种结果中的任何一种都可以被选择为基准结果。

[45] 当然，如果对某个特定组别相关系数为正的，那么它就会享受“种族的意外收获”。

[46] 这些比率表示沃尔德统计量(Wald statistic)的一种形式。

[47] 除了年龄：年龄更大的人更加可能处于更高的职业类别，或者，就方程 3.14 的估计系数来说，$\hat{\theta}_{j1} > 0$，$j = 1, 2$。

[48] 除非有特殊说明相反情况，否则本文的讨论均带有其他条件相同这一条件。

[49] 虽然在 PMT 这个类别上对居住在北部的黑人并不适用。

[50] 或者说，样本中 41.5%的所有男性，41.8%的白人男性，29.3%的黑人男性，和 35.1%的印度男性处于 PMT 职业类别中。

[51] 通过把方程里面的 $\bar{Z}_j$，视情况而定，换成 $\bar{Z}_{wj}$，$\bar{Z}_{bj}$ 和 $\bar{Z}_{Ij}$。

[52] 或者同等地利用样本白人中个人的决定变量 X_k 的均值 $\bar{X}_{wr}$ 进行计算。

[53] 应该注意的是，这一部分介绍的以模拟为基础的策略是与关于组间结果差别相关的一系列具体问题有关的。至于模拟策略的更加概括性的方法，请见 Tomz, Wittenberg and King(1999)。Tom 等人(1999)开发了一个程序 Clarify，这个程序利用随机模拟把原始的统计程序结果转换成让研究者直接感兴趣的结果。这个程序是为了在 STATA 中运行而设计的。

[54] 类似的应用也可以在假定样本中每个人都为印度人的情况下实行。

[55] 其中，α_{jr} 是白人的系数，$\alpha_{jr} + \gamma_{jr}$ 是印度人的系数。

[56] $\hat{Z}_{ij}$是用表3.3中所示的估计值联合每个人的决定变量的值计算出来的。在方程3.14中,当所有人是白人时,$IND_i = BLK_i = 0$;当所有人是黑人时,$BLK_i = 1$。

[57] 另一方面,如果$\bar{p}_j^B > \bar{p}_j^W$,则$\lambda_j^B > 1$,存在种族优势。

[58] 另一方面,如果$s_j^B > s_j^W$,则$\mu_j^B > 1$,存在总体获利。

[59] 这仅仅取决于样本比例的比,而它当然是保持不变的。

[60] 在多类别模型中,风险比对数仅仅取决于与j和k相关的系数(间的差别),与任何其他结果的系数无关。

[61] Domencich and McFadden(1996)。

[62] 这个不足之处是经济学中许多在经验上方便的函数形式共有的,比如柯布—道格拉斯(Cobb-Douglas)或者不变替代弹性(CES)(Constant Elasticity of Substitution)函数形式。

[63] 虽然必须强调,新选项的引进可能是对已有选项的分割:比如把熟练的体力劳动者/非体力劳动者分成熟练的体力劳动者和熟练的非体力劳动者两个选项。

[64] 见先前讨论的如何在引进一个新选项时重新计算概率。

[65] 在网站 http://www.indiana.edu/~statmath/stat/all/cat/giant.html 可以找到。

参考文献

AGRESTI, A (*1996*). *An introduction to categorical data analysis*. New York: John Wiley.

ALDRICH, J. H., and NELSON, F. D. (1984). Linear probability, logit, and probit models, *Quantitative Applications in the Social Sciences*, *07—045*. Beverly Hills, CA: Sage.

BEN-AKIVA, M., and LERMAN, S. (198S). *Discrete choice analysis*. London: MIT Press.

BOROOAH, V. K., and CARCACH, C. (1997). Fear and crime: Evidence from Australia. *British Journal of Criminology*, *37*, 634—656.

BOROOAH, V K. (2000). Targeting social need: Why are deprivation levels in Northern Ireland higher for Catholics than for Protestants? *Journal of Social Policy*, 29, 281—301.

BOROOAH, V K. (2001). How do employees of ethnic origin fare on the occupational ladder in Britain? *Scol/ish Journal of Political Economy*, *48*, 1—26.

BRANT, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrika*, *44*, 131—140.

CRAMER, J. (1999). Predictive performance of the binary logit model in unbalanced samples. *Journal of the Royal Statistical Society*, *Series D* (*The Statistician*), *88*, 85—94.

DEMARIS, A (1992). Logit modelling. *Quantitative Applications in the Social Sciences*, *07—086*. Newbury Park, CA: Sage.

DESAI, M., and SHAH, A(1988). An econometric approach to the measurement of poverty. *Oxford Economic Papers*, *40*, 505—522.

DOMENCICH, T. A, and McFADDEN, D. (1996). *Urban travel demand: A behavioral analysis*. Amsterdam: North-Holland.

GREENE, W. H. (1995). *LIMDEP*, *version 7. 0: User's manual*. Bellport, NY: Econometric Software.

GREENE, W. H. (2000). *Econometric analysis*. Englewood Cliffs, NJ: Prentice Hall(4th ed.,).

HAUSMAN, J., and McFADDEN, D. (1984). A specification test for the multinomial logit model. *Econometrica*, *52*, 1219—1240.

HENSHER, D. (1986). *Simultaneous estimation of hierarchical logit mode*

choice models (Working Paper No. 24). MacQuarie University, School of Economic and Financial Studies.

JORESKOG, K., and SORBOM, D. (1988). *PRELIS: A program for multivariate data screening and data summarization.* Chicago: Scientific Software Inc.

JORESKOG, K., and SORBOM, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Lawrence Erlbaum.

KAY, R., and LITTLE, S. (1986). Assessing the fit of the logistic model: A case study of children with haemolytic uraemic syndrome. *Applied Statistics*, *35*, 16—30.

LIAO, T. F. (1994). Interpreting probability models. *Quantitative Applications in the Social Sciences*, *07—101*. Newbury Park, CA: Sage.

McFADDEN, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka(Ed.), *Frontiers in econometrics.* New York: Academic Press.

MENARD, S. (1995). Applied regression analysis. *Quantitative Applications in the Social Sciences.* Thousand Oaks, CA: Sage.

NOLAN, B., and WHELAN, C. T. (1996). *Resources deprivation and poverty.* Oxford: Clarendon Press.

PIACHAUD, D. (1987). Problems in the definition of poverty. *Journal of Social Policy*, *16*, 147—164.

SCHMIDT, P., and STRAUSS, R. P. (1975). The Prediction of occupation using multiple logit models. *International Economic Review.* *13*, 471—486.

STATA(1999). *Stata reference manual release* 6. College Station, TX: Stata Press.

TOMZ, M., WITTENBERG, J., and KING, G. (1999). *CLARIFY: Software for interpreting and presenting statistical results*, *version* 1. 2. 1. Cambridge, MA: Harvard University Press. (http://gking.harvard.edu/).

TOWNSEND, P. (1979). *Poverty in the United Kingdom.* Harmondsworth: Penguin.

VEALL, M. R., and ZIMMERMANN, K. F. (1996). Pseudo-R^2 measures for some common limited dependent variable models. *Journal of Economic Surveys*, *10*, *241—260*.

译名对照表

adjacent categories logit	相邻类别 logit(模型)
base outcome	“基准”结果
best linear unbiased estimator(BLUE)	最优线性无偏估计
binary choice model	二分选择模型
binary model	二分模型
categorical variable	类别变量
Cobb-Douglas	柯布—道格拉斯(函数形式)
complementary log-log distribution	互补双对数分布
Constant Elasticity of Substitution(CES)	不变替代弹性(函数形式)
continuation ratio logit model	连续比 logit 模型
critical value	临界值
Cross-Price Elasticity of Demand	需求的交叉弹性
cumulative probability distribution	累积概率分布
cutoff parameter	临界参数
degree of freedom	自由度
discrete dependent variable	离散因变量
dummy variable	虚拟变量
error term	误差项
estimate	估计值
estimation	估计
explanatory power	解释力
generalized logit model	广义 logit 模型
goodness of fit test	拟合优度检验
Independence of Irrelevant Alternatives(IIA)	无关选项独立性
Independently and Identically Distributed(IID)	独立同分布
individual observations	个体观测值
interaction effect	交互作用
interaction variables	交互变量
intercept term	截距项
latent regression	潜在回归
latent variable	潜在变量

likelihood ratio	似然比
likelihood-ratio test	似然比检验
linear function	线性函数
log likelihood function	对数似然函数
logistically distributed	逻辑分布
log-log distribution	双对数分布
marginal effect	边际效应
maximum likelihood estimates	最大似然估计值
model specificaitn	模型设定
multinomial distribution	多项分布
multinomial logit/probit	多类别 logit/probit
multinormal integration	多重正态积分
multiple outcome model	多元结果模型
negative log-log distribution	负双对数分布
nonscalar covariance matrix	非纯量协方差矩阵
normally distributed	正态分布
null hypothesis	零假设
odds ratio	比数比
ordered logit/probit	次序 logit/probit
ordinal outcomes	有序的结果
ordinal dependent variable	有序因变量
Ordinary Least Squares(OLS)	普通最小二乘法
parallel slopes cumulative model	平行斜率累积模型
parallel slopes	平行斜率
parameter	参数
perfect prediction	理想预测
probability density function	概率密度函数
proportional-odds model	比例比数模型
qualitative choice model	定性选择模型
random utility model	随机效用模型
risk-ratio	风险比
sample statistics	样本统计量

simulation	模拟
skewed ordinal data	偏斜有序数据
standard normal variate(SNV)	标准正态变量
stereotype model	构造型模型
subsample	子样本
threshold values	临界值
unbalanced sample	不平衡样本
utility maximization	效用最大化
Wald statistic	沃尔德统计量
Weibull distribution	威布尔分布函数